

This document was generated by me, Colin Hinson, from a Crown copyright document held at R.A.F. Henlow Signals Museum. It is presented here (for free) under the Open Government Licence (O.G.L.) and this version of the document is my copyright (along with the Crown Copyright) in much the same way as a photograph would be.

The document should have been downloaded from my website <https://blunham.com/Radar>, if you downloaded it from elsewhere, please let me know (particularly if you were charged for it). You can contact me via my Genuki email page:

<https://www.genuki.org.uk/big/eng/YKS/various?recipient=colin>

You may not copy the file for onward transmission of the data nor attempt to make monetary gain by the use of these files. If you want someone else to have a copy of the file, point them at the website. It should be noted that most of the pages are identifiable as having been processed by me.

I put a lot of time into producing these files which is why you are met with the above when you open the file.

In order to generate this file, I need to:

1. Scan the pages (Epson 15000 A3 scanner)
2. Split the photographs out from the text only pages
3. Run my own software to split double pages and remove any edge marks such as punch holes
4. Run my own software to clean up the pages
5. Run my own software to set the pages to a given size and align the text correctly.
6. Run Omnipage (Ultimate) on all the pages to OCR them.
7. Generate a searchable Acrobat pdf file.

Hopefully after all that, I end up with a presentable file, though I can't guarantee there aren't errors in the OCR. If you find missing pages, pages in the wrong order, anything else wrong with the file or simply want to make a comment, please drop me a line (see above).

It is my hope that you find the file of use to you personally – I know that I would have liked to have found some of these files years ago – they would have saved me a lot of time !

Colin Hinson

In the village of Blunham, Bedfordshire.

This Signal Manual was
used by my Father,
Wing Cdr. M.E. Pulvermacher
(109974).

He was based at RAF Herlow
between February 1955
and April 1957 and was
later C.O. of 264 Signals
Unit.

Catherine Pulvermacher
(donated October 2014)

R.A.F. Signal Manual, Part II
(Radio-communication)

CONTENTS

Introduction	
Chapter I	Electricity
Chapter II	Magnetism
Chapter III	D.C. measuring instruments
Chapter IV	Dynamo-electric machinery
Chapter V	Single-phase alternating currents
Chapter VI	The transformer
Chapter VII	Electric oscillations and electro-magnetic waves
Chapter VIII	The thermionic valve
Chapter IX	Radio-telegraphic transmitters
Chapter X	Radio-telegraphic reception
Chapter XI	Amplification
Chapter XII	Radio-telephony
Chapter XIII	Radio-frequency measurements
Chapter XIV	Propagation of electro-magnetic waves
Chapter XV	Aerial arrays and transmission lines
Chapter XVI	Radio aids to navigation
Chapter XVII	
Appendix A	Tables
Appendix B	Mathematical notes

INTRODUCTION

1. The science of radio-communication depends upon the application of the theories of electricity and magnetism, and before a proper understanding of the former can be gained it is necessary to have some knowledge of the elementary principles of electrical engineering. It is impossible in the space available to give anything but the briefest of outlines and for this reason the principles dealt with are restricted, as far as possible, to those actually encountered.

2. Throughout the following pages the reader should bear in mind that the effects described are capable of demonstration and the explanations advanced are put forward as working hypotheses. It is in the attempt to form mental pictures and describe in words the causes which produce these effects that difficulty occurs, and it must be understood that an explanation may be so simplified for the sake of intelligibility that it is true only when read in its context.

3. One of the first misconceptions which must be rejected is that electricity can be produced. The term production and generation of electricity must be understood to refer to the act of causing electricity to flow, that is to say, nothing is created, but electrons are set in motion, and the flow is controlled in accordance with certain laws.

4. The principle of the conservation of energy states that energy is never created and is never destroyed. Energy may cease to be available for useful purposes, but this energy will invariably be found stored in some form or other. Thus, in a Ground Radio Station, the energy may be stored in the first place as a supply of oil fuel. This energy is liberated in the diesel engine, where it is converted partly into heat and partly into motion of the pistons and crankshaft. The latter drives an electrical generator in which the mechanical energy due to rotation is converted into electrical energy and into heat. Finally in the radio transmitter itself, some of this electrical energy is radiated into space and a portion converted into heat. At each stage of conversion, a portion of the energy is wasted, but not destroyed.

5. Radio communication then, in common with all other applications of electrical engineering, deals with the transmission and conversion of energy, which is defined as the ability to do work. If a body possesses this ability by virtue of its position it is said to possess potential energy, while if its capability is due to the fact that it is in motion it is said to possess kinetic energy. Our knowledge of the laws of nature has been developed by careful observation and experiment, and in these investigations the conceptions of time, space, mass and force are of primary importance. Those taken as fundamental are space, mass and time. Space is measured by its linear dimensions in feet, or metres, while masses can be compared by means of balances or weighing machines. The measurement of time must be derived from solar observation, but in practice is obtained from some form of clock. In all cases it is necessary to postulate some standard of comparison. The English standard of length is the imperial yard, which is deposited in the Board of Trade, and of which copies are maintained in the Royal Mint and other places. For engineering purposes the foot or $\frac{1}{3}$ yard is generally used.

6. The English unit of mass is the imperial standard pound avoirdupois, which is a piece of platinum preserved by the Board of Trade. Both the above units are arbitrary, that is they have no basis in natural phenomena. The unit of time, however, is a natural one. The sidereal day is the period of the earth's rotation on its axis, and from this is derived the mean solar day, or average duration of the sidereal day. This is divided into 24 hours, each containing 60 minutes of 60 seconds. Thus the mean solar second is $\frac{1}{86400}$ part of a mean solar day and this is the standard of time both in physical science and engineering.

7. For scientific purposes, the metric system is in use to a greater extent than the English system of units. Here the unit of time is also the mean solar second, but the unit of length is the metre, which was originally intended to be one ten-millionth part of the distance from the north pole to the equator, measured on the surface of the earth. However, in practice it is the length of a certain platinum rod which is preserved in the French archives. The metric unit of mass is the gram. This is the mass of a certain quantity of distilled water, at 4° C., but actually a platinum standard is maintained.

8. In scientific work the units of length, mass, and time are the centimetre, gram and second, and the system of measurement based on these is called the C.G.S. system. British engineers still employ the foot-pound-second or F.P.S. units for many practical purposes. The science of mechanics is based upon the three laws of motion. These laws were formulated in their present form by Isaac Newton (circa 1687) although they were previously known. They are as follows :—

First Law.—Every body continues in its state of rest or of uniform motion in a straight line, except in so far as it may be compelled by force to change that state.

From this law a definition of force is obtained.

DEF. Force is any cause which alters or tends to alter the state of a body, whether of rest or of uniform motion in a straight line.

Second Law.—Change of motion is proportional to the applied force, and takes place in the direction in which the force acts.

Third Law.—To every action there is an equal and opposite reaction.

9. The practical interpretation of the first law is that the application of force to a body will change its velocity in some way. The unit of velocity is the foot per second (F.P.S. system) or centimetre per second (C.G.S. system). A body moving with uniform velocity will pass over equal distances in equal times, no matter how small those time intervals may be. If s = the space passed over, u the uniform velocity of movement, and t the duration of the motion, these quantities are related by the equation

$$s = u t.$$

10. When the velocity of a body is changing the body is said to undergo acceleration. The acceleration is measured by the rate of change of velocity. Its unit is the "foot per second per second" (F.P.S. system) or "centimetre per second per second" (C.G.S. system). If a body is moving with a velocity of u feet per second, and is subjected to a uniform acceleration a in the direction in which it is moving, its velocity after a time t is increased from u feet per second to $u + a t$ feet per second. The second law therefore signifies that the application of force to a body results in acceleration. The acceleration is in fact proportional to the force applied, and inversely proportional to the mass of the body, hence, if f is the force, m the mass and a the acceleration produced,

$$a = \frac{f}{m} \text{ or } f = m a.$$

11. The law of gravitation asserts that every particle of the matter in the universe attracts and is attracted by every other particle. The earth therefore tends to attract other bodies, and when applied to objects near its surface this force of attraction is manifested as weight. It is important to remember that the weight of a body is the force with which it is attracted by, and mutually attracts the earth. When different bodies fall freely in a vacuum they acquire the same acceleration and therefore the force of gravity acting upon a body is proportional to its mass. For this reason the invariable practice is to compare the masses of two bodies by reference to their respective weights, and this sometimes leads to confusion.

12. By careful experiment it is found that the average acceleration due to gravity (in the British Islands) is 32.2 feet per second per second. This is usually denoted by g . The British unit of force is that which will produce in a mass of one pound, an acceleration of one foot per second per second. This unit is called the poundal. Its relation to the weight of one pound of matter is derived as under :—

1 pound-weight will produce in 1 lb. mass an acceleration of g feet per second per second.

Also 1 poundal will produce in 1 lb. mass an acceleration of 1 foot per second per second.

∴ 1 pound weight = g poundals.

Algebraically, if a mass m is allowed to fall freely its acceleration is g , the force acting is the weight w of the mass m , and

$$w = mg.$$

Example.

A mass of 30 lb. is acted upon by a force which produces in one second a velocity of 16 feet per second. Find the magnitude of the force.

$$m = 30 \text{ lb.}, a = 16 \text{ ft. per second per second.}$$

$$\begin{aligned} f &= ma \\ &= 30 \times 16 = 480 \text{ poundals.} \end{aligned}$$

For many heavy engineering purposes, however, the pound-weight (Lb.) is used as the unit of force and the unit of mass is then that mass upon which the force of 1 Lb. produces an acceleration of 1 foot per second per second. Obviously this mass is equal to g pounds of matter. The last example would then become

$$\begin{aligned} m &= \frac{30}{g}, a = 16 \text{ ft. per second per second.} \\ f &= \frac{30}{g} \times 16 = \frac{30 \times 16}{32.2} = 14.9 \text{ Lb.} \end{aligned}$$

To avoid confusion, the symbol "lb." is used to denote pound (mass) and "Lb." to denote pound (force).

13. The momentum of a body is the product of its mass and its velocity or momentum = mv . As acceleration is the rate at which velocity is varying, the force acting on a body may also be measured by the rate at which the momentum is changing. For example, take a $2\frac{1}{2}$ lb. hammer head which is moving with a velocity of 40 feet per second and is brought to rest in .001 second. The average force of the blow will be the rate at which the momentum is destroyed or $\frac{mv}{t}$.

$$\begin{aligned} f &= \frac{2.5 \times 40}{.001} \\ &= 100,000 \text{ poundals.} \end{aligned}$$

Alternatively

$$f = \frac{2.5}{32.2} \times \frac{40}{.001} = 3,105 \text{ Lb.}$$

The effect of force acting through a given space is called the work done. Both force and motion are required to perform work. A body in uniform (straight-line) motion does no work because no force is required to maintain the motion. The unit of work is the foot-pound (ft.-Lb.) and is the work done in overcoming the resistance caused by gravity on a mass of one pound.

Algebraically

$$W = fd.$$

14. It may be noted here that because this unit of work is in universal use the so-called absolute unit of force, the poundal, is hardly ever used by practical engineers.

Examples.—(i) Find the work done by gravity when a mass of 8 lb. falls from a height of 10 feet.

$$\begin{aligned} W &= fd. \\ &= 8 \times 10 = 80 \text{ foot-pounds (ft.-Lb.).} \end{aligned}$$

(ii) If the pressure on the piston of a petrol engine is 100 Lb. per square inch, the piston-head area 4 sq. in. and the length of stroke is 4 inches, find the amount of work done by a single stroke.

Total force = pressure $\times$ area

$$= 100 \frac{\text{Lb.}}{\text{in.}^2} \times 4 \text{ in.}^2$$

$$= 400 \text{ Lb.}$$

$$\text{Work} = \text{force} \times \text{distance} = 400 \text{ Lb.} \times \frac{4}{12} \text{ feet}$$

$$= 133.3 \text{ ft.-Lb.}$$

The power of an agent is the rate at which it can do work. If the above engine makes 600 strokes per minute the power of the machine is

$$\frac{400}{3} \text{ ft.-Lb.} \times \frac{600}{\text{min.}} = 80,000 \text{ ft.-Lb. per minute.}$$

The practical unit of power is the horse-power which is 33,000 ft.-Lb. per minute. Hence the above engine develops a horse-power of $\frac{80}{33}$ or 2.42 horse-power.

15. Energy has already been defined as the capability to do work, and its units are the same as those of work itself. The potential energy stored in a body of mass m , at a height h feet above the earth's surface is $m h$ ft.-Lb. or $m g h$ foot poundals. Thus if $m = 10$ lb. and $h = 144$ feet, the potential energy stored is 1,440 ft.-Lb. or 46,400 foot-poundals. If allowed to fall freely from this height, it acquires kinetic energy and the kinetic energy gained is equal to the potential energy lost. The kinetic energy possessed by a mass m moving with a velocity u is $\frac{1}{2} m u^2$ foot poundals or $\frac{1}{2} \frac{w}{g} u^2$ ft.-Lb. w being the weight of the body as before.

16. A body falling from a height of 144 feet acquires a final velocity of approximately 96 feet per second. The kinetic energy possessed by the above body at the end of its fall will be

$$\frac{1}{2} \times 10 \times 96^2 \text{ foot poundals}$$

$$\text{or } \frac{1}{2} \times \frac{10}{32.2} \times 96^2 \text{ ft.-Lb.}$$

$$= 46,100 \text{ foot poundals.}$$

$$= 1,430 \text{ ft.-Lb.}$$

Other examples of bodies possessing potential energy are a spring under tension or compression and compressed air.

17. The conservation of energy has already been mentioned. It may be asked what becomes of the kinetic energy, 1,430 ft.-Lb. possessed by the body at the instant of impact with the earth. The answer is that it is converted into yet another form of energy, namely heat. While it is very easy to transform a given amount of energy into heat, no means are known by which the whole of a given amount of heat can be converted into potential or kinetic, that is, useful energy. For this reason heat is regarded as the lowest form of energy.

18. Hitherto, most points have been illustrated with regard to the F.P.S. system. In the C.G.S. system the unit of mass is the gram and that of acceleration the centimetre per second per second. The unit of force is that which gives unit acceleration. Now the value of g in centimetres per second per second is 981, and the unit of force must be $\frac{1}{981}$ of a gram. This unit is called the dyne. Again as work = force $\times$ distance the unit of work is one dyne acting through a distance of one centimetre, and is called the dyne-centimetre or erg.

19. Electrical units are derived from the C.G.S. system. In addition to the fundamental units, which are often inconvenient for practical computation, a system of practical units has been developed. Of these the only one of immediate interest is the Joule which is a practical unit of work and is equal to 10^7 ergs.

20. To illustrate the procedure of conversion the example of the petrol engine will be reworked in C.G.S. units. The method of conversion is as follows. As

$$1 \text{ lb.} = 453 \text{ grams.}$$

$$\frac{1 \text{ lb.}}{453 \text{ grams}} = 1, \text{ or } \frac{453 \text{ grams}}{1 \text{ lb.}} = 1$$

and multiplying by a fraction of this kind will only alter the units in which the result is expressed.

The pressure on the cylinder head is $100 \frac{\text{Lb.}}{\text{in.}^2} \times 4 \text{ in.}^2$ or 400 Lb.

$$400 \text{ lb.} \times \frac{453 \text{ grams}}{1 \text{ lb.}} = 400 \times 453 \text{ grams.}$$

Hence the work done in a single stroke is

$$\begin{aligned} 400 \text{ lb.} \times \frac{453 \text{ grams}}{1 \text{ lb.}} \times \frac{981 \text{ dynes}}{1 \text{ gram}} \times 4 \text{ inches} \times \frac{2.54 \text{ cms.}}{1 \text{ inch}} \\ = 400 \times 453 \times 981 \times 4 \times 2.54 \text{ dyne-cms. or ergs.} \\ = 1.805 \times 10^9 \text{ ergs.} \\ = 180.5 \text{ joules.} \end{aligned}$$

Since 600 strokes are executed in one minute or 10 strokes per second the power exerted by the engine is 1,805 joules per second. One joule per second is also known as one watt. It is chiefly used as an electrical unit.

From the two calculations it is easy to deduce the relationship between the horse-power and the watt for

$$2.42 \text{ horse-power} = 1,805 \text{ joules.}$$

$$\therefore 1 \text{ horse-power} = 746 \text{ watts.}$$

CHAPTER I.—ELECTRICITY

MATTER

1. Before commencing the study of electromagnetic phenomena, it is necessary to acquire at least a rudimentary knowledge of the constitution of matter.

Matter may be defined as anything which occupies space and is acted upon by gravitational forces. The amount of matter contained in a body is called its mass, and the amount of matter in unit volume the density, so that the density of a homogeneous body, that is, one which is of the same nature throughout, is its mass divided by its volume. There are three states in which matter may exist—solid, liquid and gaseous. A solid is a body which tends to retain its shape, that is, it will support a stress (or force) tending to shear it. Liquids and gases are both classed as fluids. A liquid takes the shape of the vessel which holds it, and unless acted upon by other than gravitational forces, tends to maintain a level surface, while a gas takes the shape, and occupies the whole volume, of its containing vessel.

2. With a few exceptions, any substance may exist in either of the three states, according to the temperature and pressure to which it is subjected. Thus at normal atmospheric pressure water is a liquid between temperatures of 0°C . and 100°C . It becomes a solid (ice), at temperatures below 0°C ., and a gas (steam), at temperatures above 100°C . These temperatures are called the "melting point of ice" and the "boiling point of water" respectively. An increase of pressure lowers the melting point and raises the boiling point,

Molecules

3. (i) If a homogeneous body is divided into two portions, each of these has chemical and physical properties similar to those possessed by the original substance, for example, the density is unaltered. It is not possible, however, to proceed with such sub-division indefinitely, a stage being ultimately reached at which further sub-division completely changes the properties of the substance. The smallest particle into which a given material will divide, while retaining the properties of that particular material, is called its molecule. Further sub-division of the molecule is possible, but its constituents do not exhibit the same properties as the original material.

(ii) In all states of matter, the molecules are in continuous rapid motion, that is, they possess kinetic energy. In a solid, the molecules are crowded closely together, so that, although the motion is rapid, it consists of an oscillation about a mean position. The closeness of their packing results in large attractive forces between the molecules, the phenomenon being known as cohesion, and it is this cohesion which gives the solid its rigidity and enables it to withstand shearing stress. In a liquid the molecules are less closely packed, and cohesion between molecules is insufficient to prevent their movement from point to point in the material, while in a gas, there is little or no cohesive force, and consequently the molecules are free to move in all directions. The gas therefore expands and fills the whole volume of the containing vessel, the amplitude and velocity of the molecular movement depending upon the temperature of the body. The absolute zero of temperature is that at which all the molecules would be at rest, or possess no kinetic energy. The higher the temperature, the greater are the amplitude and velocity of the molecular movement, so that in solids and liquids, an increase of temperature results in an increase of linear dimensions. In the case of an enclosed gas, however, the dimensions are fixed by the containing vessel, and an increase of temperature is accompanied by an increase of pressure upon the walls of the container.

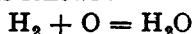
Atoms

4. Molecules are capable of sub-division, but the resulting particles are no longer molecules. They are called atoms, and have properties different from the molecules of which they formed a part, except in certain substances of which the molecule consists of only one atom. An atom is the smallest portion of matter which can enter into chemical combination, or which is obtainable by chemical separation.

CHAPTER I.—PARAS. 5-7

Elements and Compounds

5. The term element is used to denote a substance whose molecule is composed entirely of the same kind of atom, e.g. the molecule of hydrogen consists entirely of hydrogen atoms. Hydrogen is therefore an element. The complete series of elements which enter into the constitution of the universe is believed to number ninety-two, but a few of these have yet to be discovered or isolated. This belief is based upon the regularity of grouping of the known elements, and the existence of a few gaps in the regular series obtained by arranging elements in order according to the number of protons contained in the nucleus. Each element is allotted a chemical symbol, e.g. hydrogen H, oxygen O, etc. Molecules formed from atoms of different kinds are called chemical compounds, thus two atoms of hydrogen (H_2) combine with one atom of oxygen (O) to form water (H_2O). The result of such a chemical reaction is frequently exhibited in the form of a chemical equation as follows :—



Again, one atom of sodium (Na) combines with one atom of Chlorine (Cl) to form common salt or sodium chloride (Na Cl).

Atoms were for a long time thought to be absolutely indivisible—the bricks from which the whole universe was constructed. Modern research, however, has shewn that, small as an atom is, it consists of infinitely smaller particles called electrons and protons, which are so widely separated that the bulk of the atom is mainly made up of empty space.

Electrons

6. (i) An electron is an elementary particle of negative electricity. The word “is” has been used in an endeavour to emphasise that the electron has no existence whatever apart from the charge of electricity with which it is identified. This charge is generally denoted by “ e ”, and is extremely small compared with the practical unit of quantity of electricity. An electron possesses none of the ordinary properties of matter, and all electrons have identical properties irrespective of the kind of matter from which they have been derived, or with which they are associated. They have been measured and weighed by ingenious and often laborious methods, so that the radius of the electron is known to be of the order of 10^{-13} cm., and its mass 9×10^{-28} gram; this is $\frac{1}{1850}$ th of the mass of the lightest known atom, that of hydrogen. The radius of a hydrogen atom is of the order of 10^{-8} cm., about 100,000 times that of the electron. An idea of the relative magnitudes of the electron and the atom may be gained by considering a hydrogen atom magnified to the size of an ocean liner. The electron contained in it would then approximate in dimensions to the head of a pin.

(ii) Owing to their electric charge, all electrons exercise upon each other a repulsive force, which is enormous compared with their size. If two electrons were placed 1 cm. apart, the force exerted between them would be, in round figures, $\frac{1}{2.5 \times 10^{24}}$ of a pound. In case this may seem utterly insignificant, let us suppose that by some means it is possible to compress a large number of electrons into a sphere weighing one gram, or roughly $\frac{1}{450}$ of a pound. Two such spheres, placed 1 centimetre apart would repel each other with a force amounting to many millions of tons.

Protons

7. A proton is an elementary particle of positive electricity, and carries a charge equal to that of one electron. The force exerted between a proton and an electron is one of attraction, while two protons or two electrons mutually repel each other, giving rise to the first Law of Electrostatics, which may be stated in this way :—“Like charges repel and unlike charges attract each other.” The mass of a proton is very much greater than that of an electron, being 1.63×10^{-24} gram. To all intents and purposes, therefore, the mass of an atom is entirely due to that of its protons, the contribution of the electrons being absolutely negligible.

8. It is now generally accepted as a working hypothesis in electrical theory that every atom consists of a central core, or nucleus, about which one or more electrons rotate in regular orbits. The nucleus invariably contains one or more protons, together with a number of electrons, the nucleus as a whole possessing a positive charge, that is, it always contains fewer electrons than protons. In a normal atom, this positive nuclear charge is neutralised by the outer or rotating electrons. The atom thus resembles a miniature solar system, having for a sun the central nucleus, and the orbitary electrons as planets. This analogy is somewhat faulty, however, inasmuch as in the atom, the orbits of the planetary electrons are not co-planar, as is the case in our solar system. The planetary electrons are, in ordinary circumstances, retained in their orbits by the central attraction of the positively charged nucleus.

9. It has already been stated that there are believed to be ninety-two different kinds of atoms, or different elements. These atoms differ only in the number of their constituent protons and electrons. The hydrogen atom is the lightest known, having one proton only in its nucleus, and one planetary electron. The helium atom possesses four times the mass of the hydrogen atom, and is known to possess two planetary electrons. Its nucleus must therefore consist of four protons and two electrons, the latter being bound within the nucleus in some manner not yet understood. Lithium is an example of a substance having two kinds of atoms, one having six protons in the nucleus, and the other seven, so that the construction of each kind is as shewn diagrammatically in fig. 1. On the left, the constituents of the two kinds of nuclei are shewn,

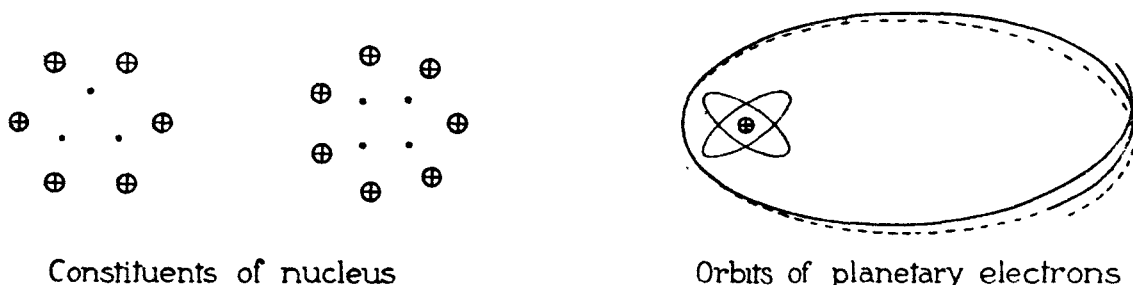


FIG. 1, CHAP. I.—Lithium atom.

protons being denoted by circles carrying a cross or positive sign, and electrons by dots. The three planetary electrons necessary to render the atom electrically neutral are believed to occupy three orbits, the two inner being circular and the outer elliptical in shape. The latter orbit may precess about the nucleus as shewn by the dotted continuation of the orbit. It must be emphasised that both kinds of lithium atoms, when entering into chemical reactions, behave identically, because both have three planetary electrons. Substances which have two forms, with different nuclear construction but the same number of planetary electrons, are called isotopic materials, each kind being called an isotope of that particular substance.

10. From the foregoing explanation it is seen that the masses of the different kinds of atoms should increase in direct proportion to the number of protons in the nucleus, and this is found to be the case. Thus the helium atom is four times as heavy as a hydrogen atom. Some lithium atoms are six and some seven times as heavy, and so on. The atomic weight of a substance is the ratio of the mass of its atom to that of the hydrogen atom. Modern research has also revealed that under certain conditions two other elementary particles may exist, namely the positive electron or positron, having a mass equal to that of the negative electron, but carrying opposite charge, and the neutron, which has a mass equal to that of a proton but is devoid of all electrical properties. Little is known about these particles owing to the rarity of their appearance, but it appears to be firmly established that they rarely if ever enter into electrical processes, and for this reason it is unnecessary to consider their properties in connection with electrical theory.

CHAPTER I.—PARAS. 11-13

Electrification

11. In certain circumstances, one or more of the planetary electrons can be detached from an atom, and is then called a free electron. Sooner or later, a free electron generally attaches itself to another atom, possibly displacing another electron in the process, which then in turn becomes free. The atom from which the electron is dislodged does not change its nature, which, as we have seen, depends essentially upon the construction of its nucleus. Atoms which have lost an electron by any means are called positive ions, and those which have gained more than their normal complement are called negative ions. The process by which atoms are caused to acquire either a surplus or deficit of electrons is called ionisation. Energy must be expended on the atom in order to detach an electron from it. If a glass rod is rubbed with a piece of silk (both being carefully dried before the experiment) it will be found that both the glass and the silk possess the property of attracting light bodies, such as dry bran or pith. The rod and silk are said to be electrified or charged, the glass positively, and the silk negatively. Actually, the glass rod has parted with some of the electrons from its superficial atoms, and these have been acquired by the silk. The results of this experiment may be summarised by stating that a positive charge is caused by a deficit of electrons in the constituent molecules of a body, and a negative charge by a surplus of electrons. Also, whenever a positive charge exists at any point, there is an equal negative charge elsewhere. That branch of the subject dealing with the phenomena associated with charged bodies is called electrostatics, the study of which will be resumed after a consideration of the action of electrons in motion.

ELECTRIC CURRENT, E.M.F. AND P.D.

12. An electric current may be defined as any movement of electrons, other than in their normal orbits within the atoms. Electric currents may be divided into three classes, depending upon the atomic or molecular mechanism by which the electronic movement takes place. They are

- (i) Conduction currents.
- (ii) Convection currents.
- (iii) Displacement currents.

Conduction Currents

13. (i) A conductor of electricity is a substance in which spontaneous ionisation takes place. Such substances usually have a large number of planetary electrons—copper, for instance, has twenty-nine—and the electrons in the outer orbits, while being attracted by their own nucleus, are also under the influence of adjacent nuclei to almost the same extent. It is therefore natural to suppose that these outer electrons are attached to one particular nucleus only by comparatively feeble bonds, and as an atom vibrates, one of its outer electrons is frequently more strongly attracted by a neighbouring nucleus than by the one about which it is nominally rotating. This electron migrates to the adjacent atom, which then momentarily possesses a surplus electron. This state of the atom, however, is unstable, and in a very short space of time one of the surplus electrons is ejected, not necessarily the intruder.

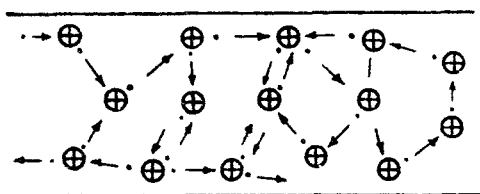
(ii) A constant transference of electrons from atom to atom is thus taking place, and at any given instant a great many free electrons are in transit from one atom to another. This migration of electrons is quite irregular, and if it were possible to count the electrons passing through any given plane in the conductor over a period of time, it would be found that just as many passed in one direction as in the opposite, so that there is no average flow in any particular direction. In certain circumstances, however, such an average flow along the conductor can be established. The flow then constitutes a conduction current. Fig. 2 is an attempt to describe an electron current pictorially, and in this figure, the direction in which each electron is moving is indicated by a small arrow.

(iii) It cannot be too strongly emphasised that, in a material carrying a conduction current, only the electrons have an average movement in one direction. The nuclei of the atoms maintain their mean relative positions with reference to each other, forming a kind of lattice, or ladder,

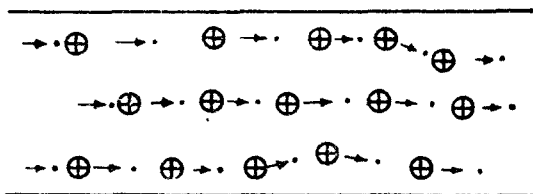
through which the electrons move, dislodging others as they pause momentarily at each atom. The actual velocity of the electrons along the conductor is very small, probably of the order of 1 centimetre per second. Pure metals are good conductors, silver being the best, and copper ranking second. Alloys, such as brass, are inferior to the pure metals in conducting property, but are still classed as good conductors.

Convection Currents

14. (i) An electrolyte is a conductor in which both positive and negative ions are free to move, and is therefore usually a liquid. Electrolytes consist of chemical compounds in solution, and in practice the term electrolyte is generally given to a solution of such a compound in water. As a concrete example, consider an aqueous solution of common salt (NaCl). The sodium atom (Na) has eleven planetary electrons arranged in three rings, the outer ring containing only one electron, while the chlorine atom (Cl) has seventeen electrons in three rings, the outer ring containing seven electrons. Now the latter arrangement is known to be somewhat unstable chemically, and a chlorine atom tends always to acquire an eighth electron on the outer orbit. When Na and Cl atoms are brought into chemical combination the sodium atom gives up its outer electron to the chlorine atom, the former becoming a positive, and the latter a negative ion, the molecule NaCl as a whole remaining electrically neutral. In the solid salt, the atoms thus united are held together by the electrical attraction between the ions. On entering into solution, however, this bond appears to be weakened, and the ions wander about between the water molecules. The name "ion" was given to these charged particles on this account, ion being the Greek word for wanderer.



No current flowing



Current flowing

FIG. 2, CHAP. I.—Conduction current.

(ii) Recombination continually occurs between the sodium and chlorine atoms, but at any given moment there are always very large numbers of free ions, an equal number of each kind being of course always in existence. If a large positive charge were introduced into one end of the vessel containing the electrolyte, and an equal negative charge into the other, a momentary electric current would be established in the electrolyte. This current would consist of negatively charged chlorine atoms moving towards the positive charge, and positively charged sodium atoms moving towards the negative charge. It is called a convection current, because it is carried by actual particles of matter, and not by electrons alone. Under certain conditions convection currents also take place in gases. The "anode current" of a thermionic valve is generally regarded as an example, although in this case free electrons play the greater part in the conduction. Nevertheless, even in the best attainable vacuum, ionised gas molecules must be present to some extent and so the term convection current is justifiable.

Displacement currents

15. In some materials, the electrons appear to be very firmly bound to their positive nuclei. Spontaneous ionisation does not occur in such substances except in a very minor degree, and they are therefore poor conductors of electricity. If ionisation never occurred at all, the substance would be an absolute non-conductor or perfect insulant. No such substance is known, and consequently there is no material which possesses perfect insulating properties. Dry air is probably the best insulant, while glass, ebonite, india-rubber, sulphur and oil are all used as

CHAPTER I.—PARA. 16

insulators, that is, to isolate a charged body or to confine an electric current to a given conductor. Although in a good insulant the number of free electrons is negligibly small, the electrons attached to atoms are acted upon by the proximity of electric charges. Let us suppose it is possible to hold a single atom of sulphur (which is an insulating material) rigidly by its nucleus, while bringing near to it a positive charge. The electrons will be attracted by the charge, and will attempt to unite with it, resulting in a displacement of the orbit of each electron, e.g. it may become elliptical and eccentric instead of circular and concentric with the nucleus. (Fig. 3.) The same effect occurs if instead of a single atom we have a body composed of insulating



FIG. 3, CHAP. I.—Displacement current.

material. The proximity of an electric charge causes the electrons to be strained from their normal orbits, towards a positive or away from a negative charge. During the very small space of time in which this motion is taking place, it constitutes a displacement current.

Constant and varying currents

16. In addition to the classification of currents according to the physical conditions under which the electronic or ionic movements take place—conduction, convection or displacement—currents are also divided into different kinds according to their variation with time. They are as follows :—

(i) *Direct current* (or D.C.).—An electric current flowing in one direction only and sensibly free from pulsation.

(ii) *Pulsating current*.—An electric current which undergoes regular recurring variations in magnitude. Both these types are referred to as unidirectional currents.

(iii) *Alternating current* (or A.C.).—An electric current which alternately reverses its direction in a circuit in a periodic manner, the frequency being independent of the constants of the circuit.

(iv) *Oscillating (or oscillatory) current*.—An electric current which alternately reverses its direction in a circuit in a periodic manner, the frequency being dependent solely on the constants of the system.

Direct currents are produced by chemical means, or by dynamo-electric machinery fitted with commutating devices or rectifying systems, while pulsating currents are found, to give only one example, in the anode circuits of thermionic valves. There is usually a slight pulsation in direct currents which are produced by rotating machinery. A pulsating current can be considered as the sum of a direct current and a series of alternating currents of different frequencies. The magnitude of an electric current of any form may be measured by means of some form of amperemeter, more shortly and commonly called an ammeter. Various types of ammeter are described in Chapter III.

Production of E.M.F.

17. If in an electrical circuit, energy of any other kind is converted into electrical energy, an electromotive force or E.M.F., is said to exist in that circuit. There are four ways of producing an E.M.F.

(i) *Chemical*, by the immersion of two dissimilar conducting substances in an electrolyte, chemical energy being transformed into electrical energy. The production of E.M.F. by this method will be dealt with in the section on primary cells.

(ii) *Thermo-electric*, by the heating of the junction between two dissimilar metals, heat energy being transformed into electrical energy. This principle is used in the service in the construction of some types of thermal ammeters which are dealt with in Chapter III.

(iii) *Frictional*, by using mechanical energy to cause friction, for example by rubbing an insulated metal cylinder with a metallic rubber, which may consist of leather coated with an amalgam of zinc and mercury. This method is of no importance as a practical method of producing an E.M.F.

It is probable that all the above methods are manifestations of one principle, i.e. that a small E.M.F. is developed whenever two dissimilar substances are in contact.

(iv) *Electromagnetic*.—The potential energy stored in a magnetic field may be combined with some kinetic energy, such as that of a moving conductor, in such a manner that some of the kinetic energy is transformed into electrical energy. This is the most practical method of obtaining electrical energy and is used in all dynamo-electric machinery.

Difference of potential

18. If between two points in an electrical circuit it is possible to convert electrical energy into any other form, a difference of potential is said to exist between the two points. The conception of difference of potential (or P.D.) is one of the most useful in practical electrical engineering. It is a matter of everyday observation that, under the action of gravitational forces, a body will fall from a high level to a lower one, losing potential energy in its passage, but acquiring kinetic energy. If the surface of the earth were perfectly smooth, without mountains or valleys, the body would possess no potential energy when lying at any point on the surface. In speaking of this aspect of gravitation, the word energy is often omitted, and the body is said to fall from a point of high potential to a point of lower potential. The word "potential" is thus synonymous with "level," and if two points are at different levels there is said to be a difference of potential between them.

Translating this notion into electrical terminology, the earth is regarded as having zero potential, and any point may then be described as "above earth potential" or "possessing positive potential with respect to earth," if the passage of a positive charge from that point to earth converts electrical energy into some other form. This can easily be visualised as the repulsion of a positive charge from the point of positive potential, resulting in mechanical movement of the charge. If the passage of a positive charge from the earth to a given point is accompanied by a conversion of electrical energy into some other form, the point is said to be below earth potential, or "possessing negative potential with respect to earth". Instruments designed for the measurement of P.D. are known as voltmeters; several types are described in Chapter III. A hydraulic analogy of P.D. is shewn in fig. 4, in which the pump may be regarded as a device converting mechanical energy into hydraulic energy, while the difference of pressure between different points on the output pipe is shewn by the difference of water level in the stand pipes. In the corresponding electrical circuit the electric battery may be regarded as a device for converting chemical into electrical energy. A difference of electrical potential exists between any two points in the circuit, owing to the resistance of the latter to the flow of electric current, and this is shewn by the difference in reading of the voltmeters. If there is a break in the electric circuit no current can flow, and there is no P.D. between any two points in the wire. This

CHAPTER I.—PARAS. 19-20

corresponds to the pipe in the water circuit being stopped up, under which conditions no water can flow, and no difference of level exists between individual stand pipes. The hydraulic analogy will be found of great assistance later, when considering the action of an electrical condenser in a circuit containing a source of alternating E.M.F.

Direction of current

19. If conduction currents only had to be considered the most natural method of defining the direction of current would be that in which the free electrons moved. We have seen, however, that a convection current consists of ions flowing in both directions, and it must be borne in mind that the chemical effects of the electric current in electrolytes were almost the first electrical phenomena to be thoroughly investigated. In this early work it was assumed that the direction of current was that of the average flow of positive ions, and this convention is still used. Thus it

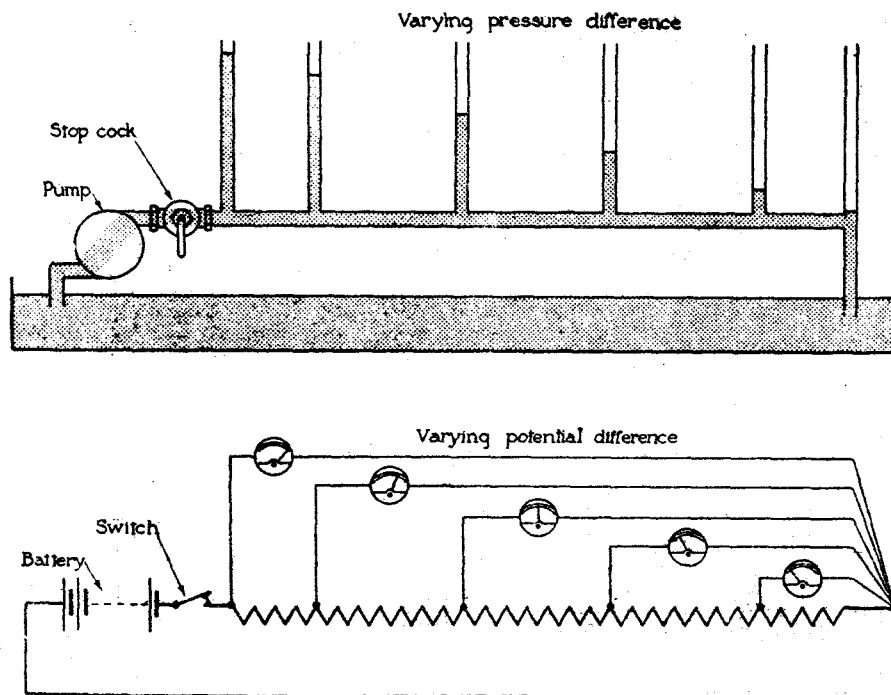


FIG. 4, CHAP. I.—Hydraulic analogue of P.D.

is considered that an electric current flows from a point of high potential to a point of lower potential, in conformity with the gravitational analogy given in the preceding section. In a complete electric circuit the current is assumed to flow in the external circuit from the positive to the negative terminals of the source of E.M.F., but through the source of E.M.F. from negative to positive terminals. When for any reason it is necessary to refer specifically to the direction of movement of the electrons, the term "electron current" will be used.

Effects of an electric current

20. When an electric current is flowing the following effects may be observed.

(i) *Heating effect.*—It has already been stated that every conversion of energy from one form to another necessitates the degradation of some portion into heat. In an electric conductor, the passage of a current results in the heating of the conductor and consequently, by radiation and convection, its immediate surroundings.

(ii) *Magnetic effect.*—The movement of an electron is always accompanied by the production of a magnetic field, and this effect is of the greatest importance in electrical work. The relation between electricity and magnetism is discussed more fully in succeeding chapters.

(iii) *Chemical effect.*—This has already been mentioned in connection with convection currents. If a current is passed through the solution of a metallic salt, pure metal is deposited upon one of the electrodes by which the current is led through the solution and gas is evolved at the other. This process is known as electrolysis.

Either of the above effects can be used to measure the rate of flow of electricity, that is, the magnitude of an electric current. The first two are commonly employed in indicating instruments, i.e. those provided with a pointer and scale graduated in units of current.

Practical electrical units

21. The deposition of metals by electrolysis lends itself admirably to the standardisation of "quantity of electricity". The practical standard of quantity is the coulomb, which is equal to 6.29×10^{18} electrons. This unit was established before the actual charge of an electron was discovered. In an electrolyte consisting of a 10 per cent aqueous solution of nitrate of silver, the passage of one coulomb invariably deposits .0011180 gram of silver at one electrode, which is called the cathode. Now mass is a physical quantity which can be determined with very great accuracy, and the coulomb can be standardised with equal precision. The practical unit of current is a rate of flow of one coulomb of electricity per second. This unit is called the ampere. The international ampere is defined as the unvarying electric current which, when passing through a 10 per cent solution of nitrate of silver in water, deposits silver at a rate of .0011180 gram per second.

The word "unvarying" is important. If the current is perfectly steady, the relationship between the ampere and the coulomb can be written algebraically

$$Q = I \times t$$

where

Q = quantity of electricity in coulombs.

I = current in amperes.

t = time during which current flows.

22. The practical unit of E.M.F. is the volt. The E.M.F. in a circuit is one volt if the amount of energy converted into the electrical form is one joule per coulomb of electricity passing. The symbol for E.M.F. is E . The relation between the energy converted W , the E.M.F. E and the quantity Q can be expressed by the equation

$$E = \frac{W}{Q}$$

The idea of energy conversion also enters into the conception of potential difference, and the volt is therefore also used as the practical unit of P.D.

If between two points in a circuit, one joule of electrical energy is converted into some other form, for each coulomb which passes from one point to the other, then the P.D. between them is one volt, i.e.

$$V = \frac{W}{Q}$$

The units of E.M.F. and P.D. having been derived, it will be observed that although we speak of electromotive force, the latter is not analogous to force in mechanics, and therefore the pump in fig. 4 is not described as "forcing water round the circuit", but as "converting mechanical into hydraulic energy". Nevertheless, the number of quarts of water moved corresponds to the quantity of electricity, and a flow of I quarts per second to a current of I coulombs per second or I amperes.

CHAPTER I.—PARAS. 23-24

Production of E.M.F. by chemical action

23. The existence of an electric current in a circuit signifies that energy in some other form is being converted into electrical energy. When chemical reactions take place energy is liberated—a most impressive example being furnished by the explosion which accompanies chemical action between the constituents of nitro-glycerine. One device by which chemical energy can be partly transformed into electrical energy is known as a primary cell.

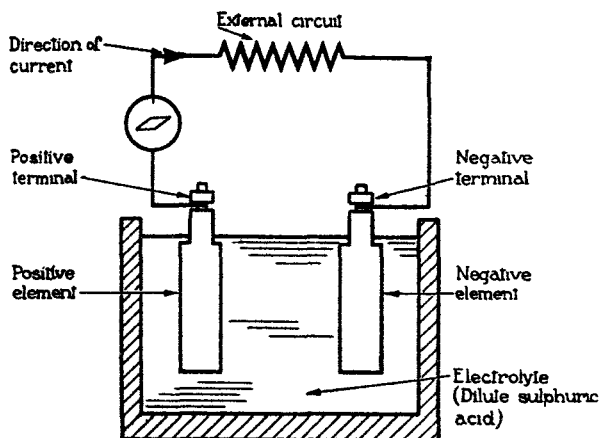


FIG. 5, CHAP. I.—Simple cell.

The simple primary cell consists of two plates, of zinc and copper respectively, immersed in a suitable electrolyte, e.g. dilute sulphuric acid. If these plates are connected externally by a conducting wire, a complete electric circuit is formed and an electric current will be established. This current is due to the electromotive force of the cell, and is accompanied by the liberation of hydrogen at the copper plate. This effect is known as polarisation. It is an undesirable phenomenon, inasmuch as it increases the resistance of the circuit, and also reduces the E.M.F. obtainable. Polarisation can be much reduced, but not entirely prevented, by supplying oxygen to the copper plate in such a way that it may combine with the hydrogen to form water. The copper plate, with the terminal attached thereto, is known as the positive element of the cell, and the zinc plate, together with its terminal, as the negative element.

24. The electrolyte used in a primary cell is not necessarily dilute sulphuric acid. Cells of the leclanche type make use of a solution of sal-ammoniac (ammonium chloride, NH_4Cl), the positive element being a carbon rod, and the negative element a zinc plate or rod. Possibly the simplest form of cell using a depolarising agent is the air depolariser cell. In this type, the positive element is in the form of a massive, thick-walled cylinder of porous carbon, while the electrolyte is a sal-ammoniac solution. The hydrogen liberated at the carbon electrode combines with oxygen occluded in the pores of the carbon, forming water. The oxygen so used is replaced by a supply from the air outside the cell by atmospheric pressure. This type of cell is not suitable for heavy currents, but it has been used in the service for intermittent supply of about .2 ampere for the filament current of a portable receiver.

The leclanche cell proper (fig. 6) makes use of the same chemical agents as the above, but in addition chemical depolarisation is resorted to, the depolariser being manganese dioxide, which is mixed with crushed carbon and packed round the positive element, the whole being contained in a porous pot of unglazed earthenware. Chemical action between the zinc and the electrolyte results in the formation of zinc chloride, ammonia gas, and hydrogen. The ammonia gas is given off at the zinc electrode, while the hydrogen passes to the positive element, where

it is oxydised by association with the manganese dioxide, forming water. This depolarisation takes place very slowly, and the leclanche cell is only suitable for small currents or for intermittent use. The E.M.F. of a leclanche cell (or of an A.D. cell) is about 1·4 volt.

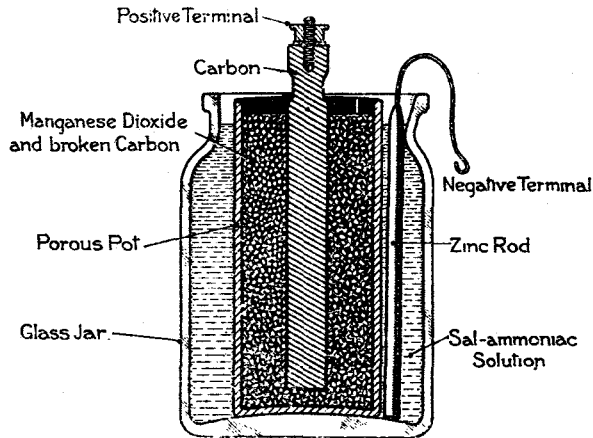


FIG. 6, CHAP. I.—Leclanche cell.

25. Dry cells are invariably of the leclanche type, but the electrolyte is in the form of a moist paste instead of a liquid solution. This paste tends to dry up with age even if the cell is unused. Batteries of dry or inert cells are used for H.T. and grid bias supplies to most service radio receivers. A typical construction is shown in fig. 7. Inert cells are leclanche type cells and

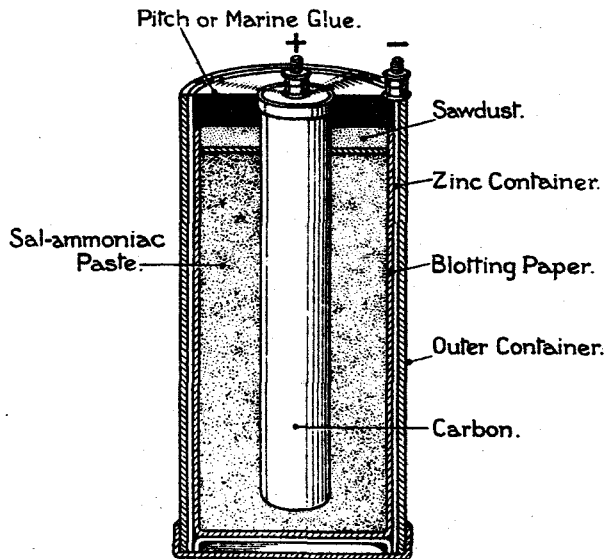


FIG. 7, CHAP. I.—Dry cell.

are similar in construction to dry cells, but are supplied with the sal-ammoniac in crystal form, needing the addition of water before the electrolyte is formed. They are therefore quite inactive until prepared for use, and therefore compare favourably with dry cells as regards shelf life, i.e. their depreciation under conditions of storage.

CHAPTER I.—PARA. 26

In a secondary cell, electrical energy is supplied and converted into chemical energy, and is available for re-conversion, into electrical energy when a suitable circuit is connected to the cell. An outline of the theory of this type of cell is contained in Air Publication 1095.

OHM'S LAW

26. Having standardised the units of current, E.M.F. and P.D., it is desirable to ascertain how these two quantities are related. To illustrate this relationship, let us take a definite conductor, say one mile of insulated electric lighting cable, and measure the current which flows when various values of E.M.F. are applied to its ends. It is assumed that the temperature of the conductor will remain constant throughout the experiment. The results may be exhibited in graphical form, the value of the current being plotted as ordinate against applied E.M.F. as abscissa, as in fig. 8. Such a graph is called the characteristic curve of the conductor, although

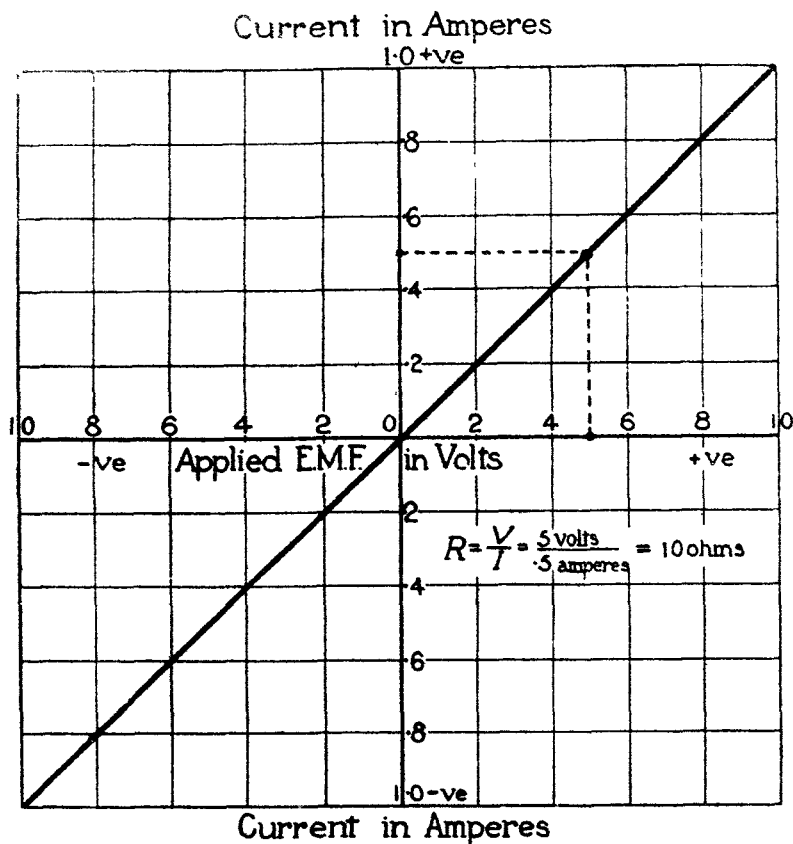


FIG. 8, CHAP. I.—Characteristic curve of conductor.

in point of fact it is a straight line through the origin of the graph. The significance of this straightness is that the current is directly proportional to the E.M.F. Expressed algebraically, $I \propto E$, or $I = G E$, G being a constant of proportion. Any conducting substance, or combination of conducting substances, which exhibits this linear relationship between current and voltage, is said to obey Ohm's law. Ohm's law is applicable either to a portion of the circuit where V is the P.D. between the two ends of that portion, or to a whole circuit. In the former case, the current is proportional to the terminal P.D., V , so that $I = G V$.

The constant G depends upon the material and dimensions of the conductor. It is known as its conductance, and is a measure of the number of free electrons (or of positive and negative ions) existing in the material. It is frequently more convenient to speak of the opposition of a conductor to the flow of an electric current, which is called its resistance, and is denoted by the symbol R . The conductance and the resistance stand in reciprocal relationship, so that $R = \frac{1}{G}$.

The unit of resistance is the ohm, while the unit of conductance is the siemens; an older name for the latter unit is the mhd. We have seen that the coulomb, and therefore the ampere, is capable of very exact standardisation by measurement of mass and time. The ohm can also be established by measurements of length and mass, and the international ohm is defined as follows:—

The international ohm is the resistance at 0°C. of a column of mercury 106.300 cms. high and 1 sq. mm. in cross section, its mass being 14.4521 grams.

27. The volt, being based theoretically upon the capability of a charge to do work, does not lead itself to standardisation by direct measurement. It is therefore established by means of Ohm's law, from the standards of current and resistance. In order to do this it is convenient to state the law in the form $V = IR$, leading to a definition of the international volt:—

The international volt is the P.D. existing between the ends of a conductor whose resistance is one international ohm, when a current of one international ampere is flowing in the conductor.

Ohm's law can only be applied to steady currents, and even then does not hold for every possible kind of conducting material. The manner in which Ohm's law is modified when dealing with varying currents is dealt with in Chapter V. Conductors which do not obey Ohm's law even for steady currents are frequently called non-ohmic conductors. There are two important classes of these.

(i) Certain combinations of metallic oxides, metallic sulphides, and metals, in contact with each other, generally referred to as crystal rectifiers.

(ii) Ionised gases carrying convection currents.

These non-ohmic conductors have important properties which will be referred to in later chapters. In particular, the second class will be dealt with at some length in Chapter VIII. In future, whenever a conductor is referred to, it will be assumed to obey Ohm's law, unless the contrary is explicitly stated.

Resistance

28. The resistance of a conductor depends upon its material, its dimensions, and its temperature. The nature of the material determines the number of free electrons, while the temperature determines the amplitude of vibration of the molecules, and therefore the degree to which free electrons are attracted by neighbouring atoms. The longer the conductor, the more certain it is that any free electron will be recaptured by an atom, while the greater the area, the more free electrons will exist in any cross section at all times. These considerations indicate, and experiment proves, that the resistance R of a conductor to an unvarying current can be expressed by the equation

$$R = \frac{l\varrho}{A}$$

where l is the length of the conductor, A its cross-sectional area, and ϱ is a constant for any particular material. The resistance thus calculated is usually referred to as the "D.C. resistance". It will be seen later that this formula must be modified if the current is varying rapidly.

Specific resistance

29. The constant ϱ is called the specific resistance or the material and may be defined either in British or metric units, depending upon the system used for the length l and cross-section A of the material. If l is in inches and A in square inches, the specific resistance is the resistance

CHAPTER I.—PARA. 29

of a sample of the material in the form of a cube each side of which is one inch in length. This standard size and shape is referred to as the "inch cube" (not the cubic inch which may have any shape whatever). If l and A are given in centimetres and square centimetres respectively the specific resistance is that of a similar sample in the form of a cube of 1 centimetre side. The resistance is measured between two opposite faces of the cube, and in either system it is convenient to refer to the specific resistance in microhms per unit cube rather than in ohms; it is also necessary to state the temperature at which the specific resistance was determined. The specific resistance of a number of common materials is given in Table I, Appendix A. The specific resistance of silver is lower than that of any other substance (1.557 microhms per centimetre cube at 15° C.) while annealed copper is a close second (1.66 microhms per centimetre cube at 15° C.).

In calculating the resistance of round wires it is convenient to state the specific resistance in ohms per mil-foot, that is the resistance of a wire one foot long and one mil (.001 inch) in diameter. The unit of area is then the circular mil, the area of a circle of diameter d mils being d^2 circular mils. This artifice avoids the introduction of the multiplier $\frac{\pi}{4}$ which would otherwise enter into the calculation of the area. To illustrate the use of the various methods the resistance of one mile of pure copper wire 0.1 inch in diameter and of uniform circular cross-section, may be calculated

(i) Given $\rho = 1.66$ microhm per centimetre cube. Both length and area must be expressed in centimetre units.

$$\begin{aligned}\text{Area} &= \frac{\pi d^2}{4} \text{ in.}^2 \\ &= \frac{\pi d^2}{4} \text{ in.}^2 \times \frac{(2.54 \text{ centimetres})^2}{1 \text{ in.}^2} \\ &= \frac{\pi}{4} \times (0.1)^2 \times 6.45 \text{ cm}^2 \\ &= .7854 \times .0645 \\ &= .05 \text{ cm}^2 \\ \text{length} &= 1 \text{ mile} \times \frac{5,280 \text{ feet}}{1 \text{ mile}} \times \frac{30.5 \text{ cm}}{1 \text{ foot}} \\ &= 5,280 \times 30.5 \text{ cms} \\ &= 161,200 \text{ centimetres} \\ R &= \frac{l \rho}{A} \\ &= \frac{161,200}{.05} \times \frac{1.66}{10^6} \\ &= 5.37 \text{ ohms}\end{aligned}$$

(ii) Given $\rho = .654$ microhms per inch cube

$$\begin{aligned}\text{area} &= \frac{\pi d^2}{4} = .7854 \times .01 \text{ in.}^2 \\ \text{length} &= 5,280 \times 12 \text{ inches} \\ R &= \frac{5,280 \times 12 \times .654}{.7854 \times .01 \times 10^6} \\ &= 5.37 \text{ ohms}\end{aligned}$$

(iii) Given $\rho = 10.18$ ohms per mil-foot
 $d = 0.1$ inch = 100 mils
 $A = d^2 = 10^4$ circular mils
 $l = 5,280$ feet
 $R = \frac{5,280 \times 10.18}{10^4}$
 $= 5.375$ ohms

Note the comparative simplicity of the third method.

Temperature coefficient

30. The effect of an increase of temperature is to increase the specific resistance of all pure metals. For all practical purposes the increase is directly proportional to the rise in temperature. If ρ_1 is the specific resistance at temperature t_1 and ρ_2 the specific resistance at temperature t_2

$$\rho_2 = \rho_1 \{1 + \alpha (t_2 - t_1)\}$$

$t_2 - t_1$ being the increase in temperature.

The constant α is called the temperature coefficient of the material and is given in Appendix A for a standard temperature of $15^\circ \text{C.} = 59^\circ \text{F.}$

Example :—

If the conductor previously considered is heated by the passage of current to 30°C. , what is its resistance ?

The specific resistance at 15°C. is $\rho_1 = 1.66$ microhms per cm. cube, and the temperature coefficient .0041.

Temperature rise = 15°C.

$$\begin{aligned}\rho_2 &= \rho_1 (1 + \alpha \times 15) \\ &= \rho_1 (1 + 0.06)\end{aligned}$$

$$\frac{\rho_2}{\rho_1} = 1.06$$

$$\therefore R_2 = 1.06 R_1 = 5.7 \text{ ohms.}$$

The temperature coefficients of alloys are very much smaller than those of pure metals while certain substances, notably carbon, have negative coefficients, that is, the resistance decreases with temperature.

ELECTRIC CIRCUIT CALCULATIONS

Kirchoff's Laws

31. These important laws are simply extensions of the idea contained in Ohm's law, and are as follows :—

First law.—At any junction of resistance the sum of the currents flowing to the junction is equal to the sum of the currents flowing away from it. This may be more shortly expressed by the statement "at any point in a circuit, the algebraic sum of the currents is zero," or more briefly still by $\Sigma I = 0$.

Second law.—In any closed circuit, the algebraic sum of the E.M.F.'s is equal to the algebraic sum of the P.D.'s. As $V = IR$, this may be written $E = \Sigma IR$, where the symbol Σ means "the algebraic sum of all such quantities as"

Circuit diagrams

32. In making diagrams of connections in electrical engineering it is necessary to employ symbols to denote the various machines and devices used. The connecting wires are shown by

CHAPTER I.—PARA. 33

lines and the points where they make electrical contact with the machine may be indicated in the diagram. In selecting and devising these symbols stress has been laid on the following points :—

Each symbol should be, as far as possible :—

- (i) Self explanatory.
- (ii) Easy to draw.
- (iii) In general use.
- (iv) Exclusive to one particular machine or device.

These symbols will be introduced to the reader by the text, as and when found desirable

Up to the present, we have only encountered two appliances which bear standard symbols, viz. chemical sources of E.M.F. and resistances, the appropriate symbols being embodied in fig. 4.

Conductors in series

33. Conductors are said to be in series when they are connected end to end, in such a manner that the same current flows through each. Each conductor must have resistance, and the circuit diagram of a number of resistive conductors in series with a primary battery is shewn in fig. 9. It must be remembered that the battery itself has resistance, although it is not

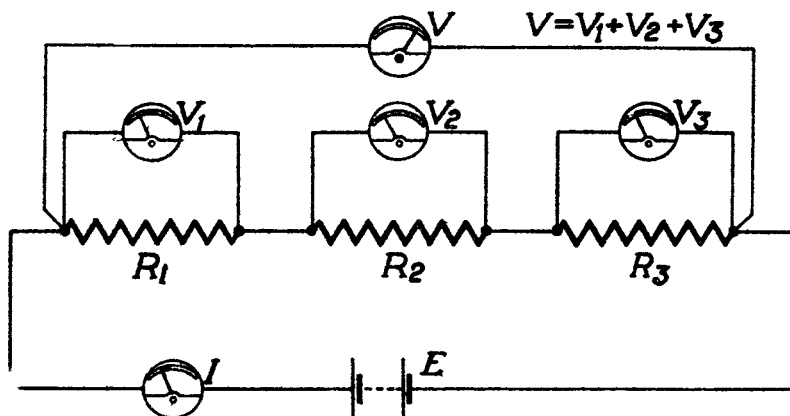


FIG. 9, CHAP. I.—Conductors in series.

customary to draw any separate symbol to shew this, but the resistance in ohms may be written beside the battery symbol. An expression for the total resistance of this circuit, containing resistances R_1 , R_2 , R_3 , will now be derived. Since there is no accumulation of electricity at any point, the same current flows through each resistance in accordance with Kirchoff's first law. Let this total current be I amps. Then by Ohm's law, the E.M.F. in the circuit is given by

$$E = IR \quad (a)$$

where R is the total resistance.

Now consider the P.D.'s between different points in the circuit

Across R_1 the P.D. is $IR_1 = V_1$

Across R_2 the P.D. is $IR_2 = V_2$

Across R_3 the P.D. is $IR_3 = V_3$

By Kirchoff's second law, $E = V_1 + V_2 + V_3$,

or $E = I(R_1 + R_2 + R_3) \quad (b)$

Comparing expressions (a) and (b), it is seen that both can be true only if

$$R = R_1 + R_2 + R_3$$

Hence the "rule for resistance in series" :—

The total resistance of a number of conductors in series is equal to the sum of their individual resistances.

Conductors in parallel

34. Conductors are said to be in parallel when they have a common P.D. between their ends, as in the diagram, fig. 10. The current in each conductor will not be equal, unless the resistance of each conductor is the same.

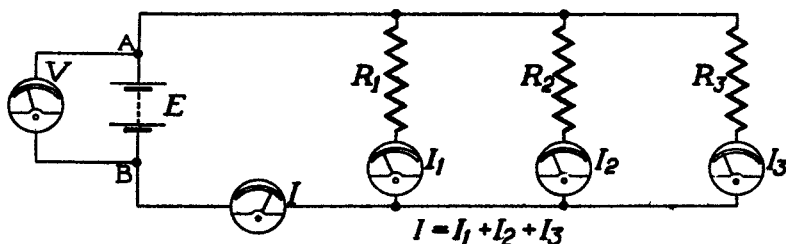


FIG. 10, CHAP. I.—Conductors in parallel.

Referring to the diagram, let the resistances of the three conductors be R_1 , R_2 , R_3 , and the corresponding currents through them be I_1 , I_2 , I_3 . By Kirchoff's first law, the current I , flowing to the point B is equal to $I_1 + I_2 + I_3$, and the latter also combine to form a current I flowing away from the point A.

That is $I = I_1 + I_2 + I_3$.

The P.D. (V) between the points A and B is equal to $I_1 R_1$ or to $I_2 R_2$ or to $I_3 R_3$.

That is

$$I_1 = \frac{V}{R_1}$$

$$I_2 = \frac{V}{R_2}$$

$$I_3 = \frac{V}{R_3}$$

$$I_1 + I_2 + I_3 = \frac{V}{R_1} + \frac{V}{R_2} + \frac{V}{R_3}$$

$$I = V \left(\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} \right)$$

but by Ohm's law $I = \frac{V}{R}$

where R is the joint resistance of the three resistances in parallel.

$$\therefore \frac{1}{R} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3}$$

and

$$R = 1 \div \left(\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} \right)$$

Note that R is less than either R_1 , R_2 , or R_3 alone. The "rule for resistances in parallel" is :—The sum of the reciprocals of the individual resistances is equal to the reciprocal of the joint resistance.

CHAPTER I.—PARA. 35

Series parallel circuits

35. When a circuit consists of combinations of conductors in series and in parallel, the procedure is to first attack all parallel combinations, reducing each to a single equivalent resistance. The various equivalent resistances can then be treated as series resistances, as in the following example :—

In the circuit given in fig. 11 find

- (i) the total resistance of the circuit,
- (ii) the current given by the battery,
- (iii) the P.D. at the terminals of the battery. N.B. this must not be confused with the E.M.F. of the battery which is 10 volts,
- (iv) the current I_s in the 3 ohm wire of branch A.

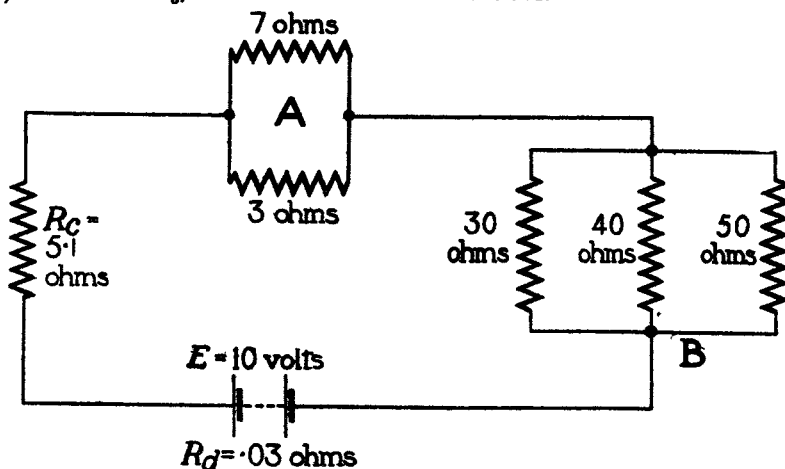


FIG. 11, CHAP. I.—Conductors in series and parallel.

(i) The method of attack is as follows. First resolve the parallel element A into an equivalent single resistance R_a .

$$\frac{1}{R_a} = \frac{1}{7} + \frac{1}{3} = \frac{3+7}{21} = \frac{10}{21}$$

$$R_a = 2.1 \text{ ohms.}$$

Next resolve the parallel element B in a similar manner.

$$\frac{1}{R_b} = \frac{1}{30} + \frac{1}{40} + \frac{1}{50}$$

$$= \frac{20 + 15 + 12}{600}$$

$$= \frac{47}{600}$$

$$R_b = \frac{600}{47} = 12.77 \text{ ohms}$$

The total resistance R is $R_a + R_b + R_c + R_d$

$$\text{or } 5.1 + 2.1 + 12.77 + 0.03 = 20 \text{ ohms}$$

(ii) The current given by the battery is $\frac{E}{R} = \frac{10}{20}$ or 0.5 ampere.

(iii) The P.D. at the terminals of the battery is equal to the product of current and resistance in the external circuit, or $I \times (R_a + R_b + R_c)$. This is 0.5×19.97 or 9.985 volts. In some examples it may be more convenient to find the terminal P.D. by remembering that it is equal

to the E.M.F. minus the “ IR drop” in the battery itself. The latter quantity is $.03 \times 0.5 = .015$ and the terminal P.D. is $20 - .015 = 9.985$ volts.

(iv) The current in the 3-ohm wire may be found by either of two methods.

(a) Find the P.D. V_a between the ends of the loop A. This is IR_a , and $I = 0.5$, $R_a = 2.1$
 $\therefore V_a = 1.05$ volts.

The current in the 3-ohm wire is then $\frac{V_a}{3}$ or 0.35 ampere.

This may be checked by finding the current in the 7-ohm wire. This is $\frac{V_a}{7}$ or 0.15 ampere. The sum of these currents is 0.5 ampere, the total current flowing.

(b) The current in parallel paths divides in inverse ratio to the resistance of each path. In the parallel element A, the 3-ohm wire carries $\left(\frac{7}{3+7}\right)$ ths and the 7-ohm wire $\left(\frac{3}{3+7}\right)$ ths of the total current. Hence $I_3 = \frac{7}{10}$ of $0.5 = .35$ ampere as already calculated by method (a).

As a further example of the manner in which current divides between different parallel paths find the currents in each of the resistances forming the parallel element B, using the second method of calculation.

Let the 30 ohms resistance be R_1 and its current $I_1 = \frac{V_b}{R_1}$

Let the 40 ohms resistance be R_2 and its current $I_2 = \frac{V_b}{R_2}$

Let the 50 ohms resistance be R_3 and its current $I_3 = \frac{V_b}{R_3}$

the P.D. between the ends of the branch being V_b . The total current I is

$$\frac{V_b}{R_b} = V \left(\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} \right)$$

$$\frac{I_1}{I} = \frac{\frac{V_b}{R_1}}{V_b \left(\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} \right)} = \frac{\frac{1}{R_1}}{\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3}}$$

$$\text{and therefore } I_1 = \frac{\frac{1}{R_1}}{\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3}} \times I$$

$$= \frac{\frac{1}{30}}{\frac{1}{30} + \frac{1}{40} + \frac{1}{50}} \times 0.5$$

$$= \frac{\frac{20}{600}}{\frac{20+15+12}{600}} \times 0.5$$

$$= \frac{20}{47} \times 0.5$$

$$= \frac{10}{47} \text{ amperes.}$$

By the same reasoning $I_2 = \frac{15}{94}$ amperes and $I_3 = \frac{12}{94}$ amperes.

CHAPTER I.—PARA. 36

Networks

36. In some circumstances the solution of a problem is facilitated by the direct application of Kirchoff's laws. An example of this type of circuit is the distribution network shewn in fig. 12

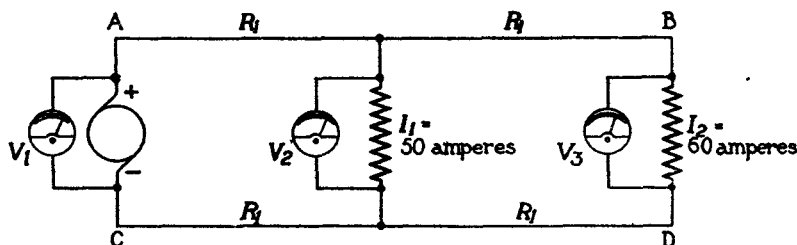


FIG. 12, CHAP. I.—Distribution network.

in which I_1 and I_2 represent two concentrated "loads" supplied from the power mains AB, CD. The P.D. at the terminals of the generator (V_1) being given it is often necessary to calculate the P.D. at the terminals of each load. The method of solving this problem is as follows.

$$V_2 = V_1 - (\text{the "IR drop" in the leads between } V_1 \text{ and } V_2) = V_1 - 2R_l(I_1 + I_2).$$

As all the quantities on the right-hand side are known, V_2 can be determined.

As an example take $V_1 = 220$ volts, $I_1 = 50$ amperes, $I_2 = 60$ amperes, $R_l = .01$ ohm.

$$\text{Then } V_2 = 220 - .02 \times 110$$

$$= 220 - 2.2$$

$$V_2 = 217.8 \text{ volts.}$$

Also

$$V_3 = V_2 - 2R_l I_2$$

$$= 217.8 - .02 \times 60$$

$$= 217.8 - 1.2$$

$$= 216.6 \text{ volts.}$$

It will be noticed that the P.D. across the load farthest from the feeding point is 3.4 volts below the P.D. at the generator end of the line. If a second generator of the same terminal P.D. were connected to the points BD the IR drop in the leads would be reduced. The distribution of the current would then be as in fig. 13 in which the actual currents supplied by each generator

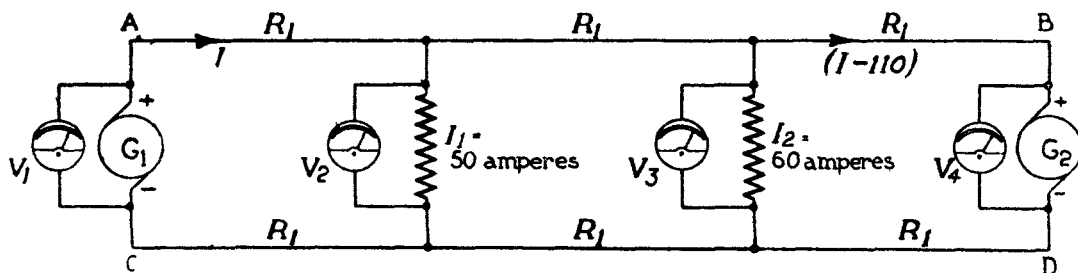


FIG. 13, CHAP. I.—Distribution network with two generators.

are unknown. It is assumed that a current I is flowing into the upper line from the terminal A, and a current $(I - \text{the total load current})$ is flowing out of the upper line into the generator G_2 at the terminal B. In the solution we shall find that the quantity $(I - \text{total load current})$ is a negative quantity, signifying that a current is actually flowing into the upper line from the generator G_2 . The following equations can be set up from the data

$$\begin{aligned}
 V_2 &= V_1 - 2R_1 I \\
 V_3 &= V_2 - 2R_1 (I - 50) \\
 V_4 &= V_3 - 2R_1 (I - 110) \\
 \text{i.e. } V_4 &= V_2 - 2R_1 (I - 50) - 2R_1 (I - 110) \\
 \text{or } V_4 &= V_1 - 2R_1 I - 2R_1 (I - 50) - 2R_1 (I - 110). \\
 V_4 &= V_1 - 2R_1 I - 2R_1 I + 100R_1 - 2R_1 I + 220R_1. \\
 \text{but } V_4 &= V_1 \\
 \therefore 0 &= -3(2R_1 I) + 320R_1. \\
 6I &= 320 \\
 I &= \frac{320}{6} = 53\frac{1}{3} \text{ amperes.}
 \end{aligned}$$

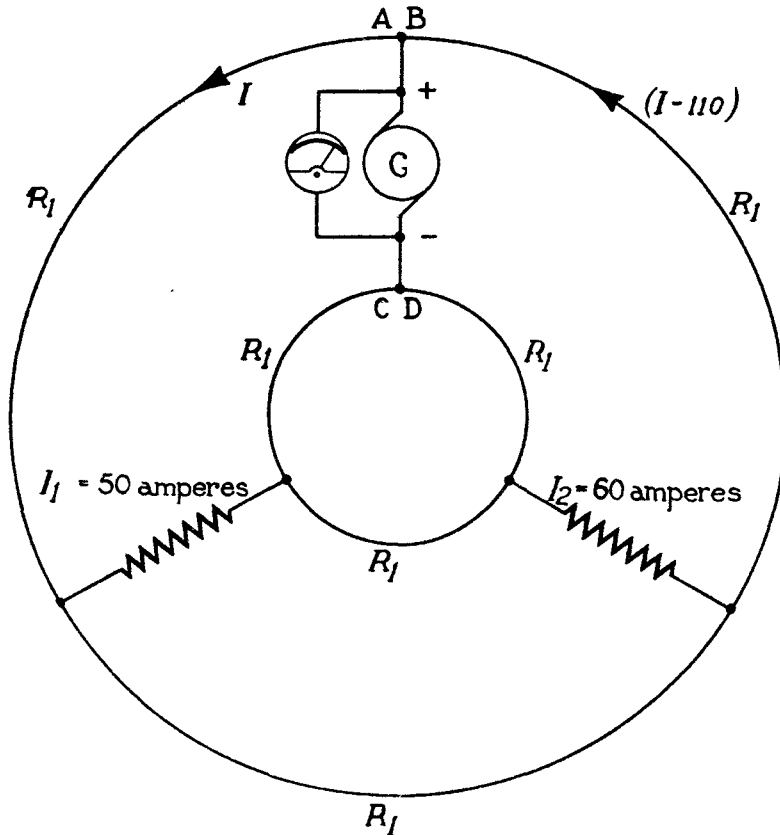


FIG. 14 CHAP. I.—Ring main system of distribution.

The current flowing into the generator G_2 at B is $(53\frac{1}{3} - 110) = -56\frac{2}{3}$ amperes, i.e. a current of $56\frac{2}{3}$ amperes flows into the upper line at this point.

CHAPTER I.—PARA. 37

The P.D's V_2 and V_3 can now be found, for

$$V_2 = V_1 - 2R_1 I$$

$$= 220 - .02 \times 53 \frac{1}{3}$$

$$= 220 - \frac{3.2}{3}$$

$$= 220 - 1.066$$

$$= \underline{218.933 \text{ volts.}}$$

$$V_3 = V_2 - 2R_1 (I - 50)$$

$$= 218.933 - .02 \times 3 \frac{1}{3}$$

$$= 218.933 - .066$$

$$= \underline{218.867 \text{ volts.}}$$

It is seen from the above that by feeding from both ends of the line the P.D. across the load is maintained at a value more nearly at the generator terminals. This principle is applied in the ring main system in which each supply main is closed upon itself, so that the points A and B are coincident, as also are C and D. A single generator then feeds the network, current flowing in either direction round the ring according to the distribution of the load (fig. 14).

Maxwell's rule

37. This is an application of Kirchoff's laws, which is often of considerable assistance in solving network problems. Consider the circuit in fig. 15 in which it is required to find the P.D. at the terminals of the resistance R , to which current is being supplied by the two batteries of E.M.F. E_1 and E_2 respectively. The rule is as follows. In every closed "mesh" of a network, consider a current to flow in clockwise direction, apply Kirchoff's law $E = \Sigma IR$ to each mesh of the network, and solve the simultaneous equations thus obtained. In the diagram, the current x

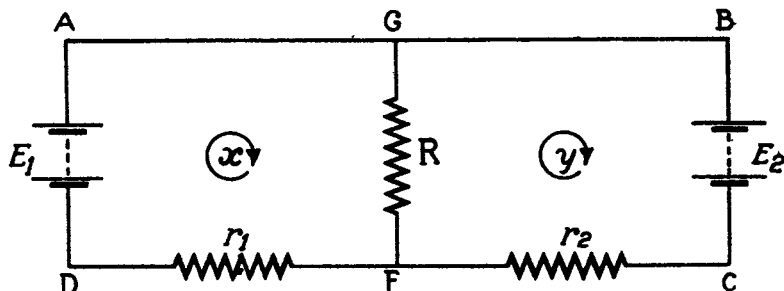


FIG. 15, CHAP. I.—Example of Maxwell's Rule.

is assumed to circulate in a clockwise direction in the mesh A G F D, while the current y is assumed to circulate in a clockwise direction in the mesh G B C F. The fact that this current is contrary to the polarity of the E.M.F. E_2 is immaterial. If y is found to be a positive quantity this will signify that the E.M.F. E_1 is sufficient to force a current to flow in the direction assumed, while if E_2 is supplying current to the resistance R , y will be found to have a negative value, i.e. its direction is opposite to that postulated in the equations.

The current through the resistance R will be $x - y$, and its terminal P.D. $R(x - y)$.

Forming the equations from the data, we have

$$E_1 = r_1 x + R x - R y$$

because both x and y flow through R and the IR drop caused by the current y must be taken into account.

Similarly

$$-E_2 = r_2 y + Ry - Rx$$

E_2 is given a negative sign because it is acting in the opposite direction to that assumed to be positive, i.e. against the current y .

These equations may be arranged thus:—

$$(r_1 + R)x - Ry = E_1 \quad \dots \quad (a)$$

$$-Rx + (r_2 + R)y = -E_2 \quad \dots \quad (b)$$

$$R(r_1 + R)x - R^2y = RE_1 \quad \dots \quad (c)$$

$$-R(r_1 + R)x + (r_1 + R)(r_2 + R)y = -(r_1 + R)E_2 \quad \dots \quad (d)$$

Equation (c) is obtained from equation (a) by multiplying all terms by R , while equation (d) is obtained from equation (b) by multiplying all terms by $(r_1 + R)$. Adding (c) and (d)

$$\begin{aligned} \{(r_1 + R)(r_2 + R) - R^2\}y &= RE_1 - (r_1 + R)E_2 \\ y &= \frac{RE_1 - (r_1 + R)E_2}{r_1r_2 + r_1R + r_2R} \quad \dots \quad (e) \end{aligned}$$

Instead of eliminating x from the equations, we may eliminate y ; if equation (a) is multiplied by $(r_2 + R)$ and equation (b) by R , we obtain

$$(r_2 + R)(r_1 + R)x - R(r_2 + R)y = E_1(r_2 + R) \quad \dots \quad (f)$$

$$-R^2x + R(r_2 + R)y = -E_2R \quad \dots \quad (g)$$

Adding these equations,

$$\begin{aligned} \{(r_2 + R)(r_1 + R) - R^2\}x &= (r_2 + R)E_1 - RE_2 \\ \text{or } x &= \frac{(r_2 + R)E_1 - RE_2}{r_1r_2 + r_1R + r_2R} \quad \dots \quad (h) \end{aligned}$$

The current through the resistance R is $x - y$

$$\begin{aligned} \text{or } \frac{r_2E_1 + RE_1 - RE_2 - RE_1 + r_1E_2 + RE_2}{r_1r_2 + r_1R + r_2R} \\ = \frac{r_2E_1 + r_1E_2}{r_1r_2 + r_1R + r_2R} \quad \dots \quad (i) \end{aligned}$$

and the P.D. between the points G and F is

$$R \frac{r_2E_1 + r_1E_2}{r_1r_2 + r_1R + r_2R} \quad \dots \quad (j)$$

As an example, suppose that $E_1 = 12$ volts, $r_1 = .1$ ohm, $E_2 = 6$ volts, $r_2 = 1$ ohm, while $R = 5$ ohms.

The P.D. at G.F. is then

$$\begin{aligned} &5 \frac{1 \times 12 + .1 \times 6}{1 \times .1 + .1 \times 5 + 1 \times 5} \\ &= 5 \frac{12.6}{.1 + .5 + 5} \\ &= \underline{11.25 \text{ volts.}} \end{aligned}$$

CHAPTER I.—PARA. 38

The direction in which this P.D. is acting is such that a current y will flow in a clockwise direction in opposition to the E.M.F. E_2 . Its magnitude will be $\frac{11.25 - E_2}{r_2} = 5.25$ amperes; this may be verified by direct calculation using equation (e) above.

Now calculate the P.D. if the resistance $r_1 = 1$ ohm and $r_2 = .1$ ohm. This is given by

$$5 \frac{.1 \times 12 + 1 \times 6}{5.6} = \frac{7.2 \times 5}{5.6} = 6.43 \text{ volts.}$$

This example shews that in certain conditions when cells or batteries are connected in parallel, it is possible for one cell or battery to force a reverse current through the other. Referring to equation (e) it is seen that the current y will be zero if $RE_1 = (r_1 + R) E_2$.

If E_1 , E_2 and r_1 are fixed, and R is varied, y will be zero if $R = \frac{r_1 E_2}{E_1 - E_2}$.

If E_1 , E_2 and r_1 , r_2 have the values given, y becomes zero if R is equal to r_1 , or 1 ohm. If R is greater than this, as in the above example, y is a positive quantity, signifying that the assumed direction of current is the correct one, or that the voltage E_1 is forcing a "reverse current" through the other battery E_2 . This might of course have been anticipated from the difference in the respective values of E_1 and E_2 , but it is not obvious, without calculation, that if R has a value less than $\frac{r_1 E_2}{E_1 - E_2}$ (i.e. less than 1 ohm in the given example) both batteries will contribute to a flow of current through the resistance R .

Although of course two batteries of unequal E.M.F. are never deliberately connected in the above manner the principle has an important application. If one of the sources of E.M.F. in fig. 15 is a secondary battery and the other a shunt-wound generator (*see* Chapter IV), the E.M.F. of the generator and battery being nominally the same, the battery is said to be "floating." So long as the equality of E.M.F. is maintained, the generator and battery will each contribute an equal current which will flow in the load circuit R , and the battery tends to discharge, that is, its E.M.F. falls as the energy stored in chemical form is converted into electrical energy. When the battery E.M.F. falls below the P.D. at the load terminals, the generator supplies a charging current to the battery restoring its E.M.F. to its normal value, while if for any reason the generator voltage fails, the battery maintains the desired current through the load resistance R . Some device is obviously required to disconnect the generator from the load in case of complete failure of its E.M.F., and this is provided by an automatic electromagnetic cutout. The operation of such devices is dealt with in A.P.1095, Electrical Equipment Manual.

Wheatstone bridge

38. A type of network which is frequently used in electrical measurements is shown in its simplest form in fig. 16, the arrangement being known as Wheatstone bridge. In the diagram, R_1 and R_2 are known resistances of fixed value, R_3 is an adjustable resistance whose value is accurately known, and R_4 is a resistance of unknown value. The object of the bridge is to measure the resistance of R_4 . In order to achieve this, a current-indicating instrument such as the simple galvanometer described in the following chapter is connected to the points A and B while a battery and switch are connected to points C and D. The resistances R_1 , R_2 , R_3 , R_4 are called the arms of the bridge while the connections between A and B, and between C and D are referred to as the diagonals of the bridge. First, assume the galvanometer circuit to be broken by the key K_2 . On pressing the key K_1 , a current will be established through the two parallel paths $R_1 + R_3$ and $R_2 + R_4$, and the current in these two branches may be denoted by I_1 and I_2 respectively. The P.D. between the ends of R_1 , i.e. between A and C, will be $I_1 R_1$, while the P.D. between B and C will be $I_2 R_2$. If these two differences of potential are equal, no difference of potential will exist between A and B, and if the galvanometer is connected to them by pressing

the key K_2 , no current will flow through it, for the addition of the galvanometer to the circuit will not alter the value of either I_1 or I_2 . The condition under which no deflection of the galvanometer will occur is therefore

$$I_1 R_1 = I_2 R_2$$

but

$$I_1 = \frac{E}{R_1 + R_3}, I_2 = \frac{E}{R_2 + R_4}$$

the condition may therefore be expressed as

$$\frac{E R_1}{R_1 + R_3} = \frac{E R_2}{R_2 + R_4}$$

or as

$$\frac{R_1 + R_3}{R_1} = \frac{R_2 + R_4}{R_2}$$

$$1 + \frac{R_3}{R_1} = 1 + \frac{R_4}{R_2}$$

or finally as

$$\frac{R_1}{R_3} = \frac{R_2}{R_4}$$

this is known as the condition in which the bridge is balanced. If as above stated R_1 and R_2 are of fixed value and the ratio between them is known, then R_3 and R_4 are in the same ratio because the last equation can be transposed giving

$$\frac{R_1}{R_2} = \frac{R_3}{R_4}$$

Thus if $\frac{R_1}{R_2} = 10$, and balance is obtained when $R_3 = 1$ ohm, then $R_4 = \frac{1}{10}$ ohm, while if $\frac{R_1}{R_2} = \frac{1}{100}$ and balance is obtained when $R_3 = 10,000$ ohms, $R_4 = 10,000 \times 100 = 1,000,000$ ohms.

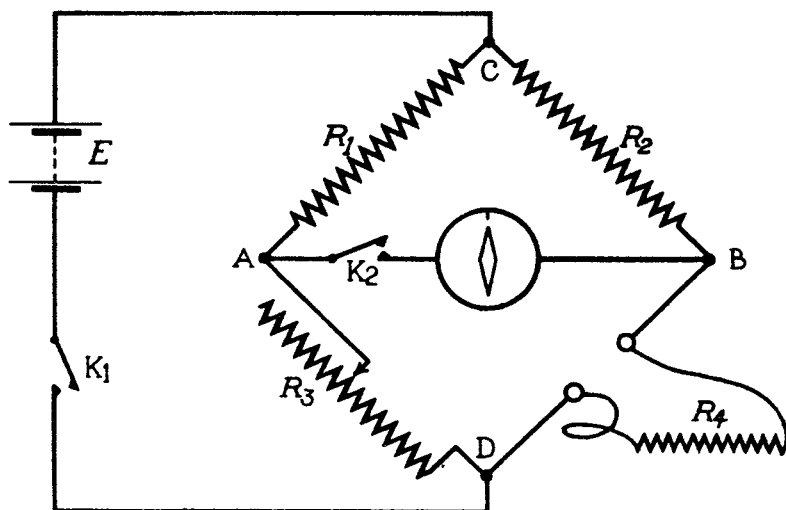


FIG. 16, CHAP. I.—Wheatstone's bridge.

CHAPTER I.—PARA. 39

The resistances R_1 and R_2 are each usually adjustable in three steps, e.g., 10, 100, 1,000 ohms, while R_3 consists of a resistance which is adjustable in steps of one ohm from zero to 9,999 ohms. The latter arrangement is called a decade resistance, and consists of four separate resistances, each of which carries 10 tapings. Calling these R_a , R_b , R_c and R_d , we have R_a adjustable in one-ohm steps, from 0 to 9 ohms, R_b in ten-ohm steps from 0 to 90 ohms, R_c in one hundred-ohm steps from 0 to 900 ohms and R_d in one thousand-ohm steps from 0 to 9,000 ohms. Any integral value of resistance from 0 to 9,999 ohms is therefore available. The principle of the Wheatstone bridge can be also applied to the measurement of capacitance and inductance.

Power

39. So far we have dealt with the conversion of energy between different points without reference to the time taken for this conversion to take place. The rate at which energy is being converted is known as power. In other words, power is the rate of doing work. The C.G.S. unit of power is the erg per second, and the practical unit is the joule per second or watt. Since one joule of energy is converted when one coulomb passes between two points whose P.D. is one volt, the power is one watt if this conversion takes place in one second. But a rate of flow of one coulomb per second is one ampere, so that the power developed between two points whose P.D. is one volt, when a current of one ampere flows, is one watt. A larger unit of 1,000 watts is also used and is called the kilowatt. (KW.)

The symbol for power is P

$$P = \frac{W}{t} = \frac{QV}{t}$$

$$\text{Since } \frac{Q}{t} = I$$

$$P = IV$$

A power of one watt developed continuously for one hour, will correspond with a transformation of $\frac{1 \text{ joule}}{\text{sec.}} \times 3,600 \text{ secs.}$ or 3,600 joules. This unit is called the watt-hour. A still larger unit, the kilo-watt-hour, (Kwh.) or Board of Trade Unit (B.O.T.U.) is also used. It is equal to 3.6×10^6 joules, and constitutes the ordinary commercial unit of electrical energy. Expressions such as "a forty-watt lamp" are occasionally used. This must be interpreted as meaning that when used at the voltage for which it was designed, it will take such a current from the supply source that the power dissipated will be 40 watts. If used on an incorrect voltage, its energy consumption will not be at the rate of 40 watts and it is desirable to state the voltage at which the appliance should be used, e.g., a 200-watt 220-volt soldering iron.

By the application of Ohm's law, expressions for power may be obtained in terms of any two of the quantities voltage, current, and resistance, for since $P = IV$, and $V = IR$.

$$(i) P = I \times IR = I^2 R$$

$$(ii) P = V \times \frac{V}{R} = \frac{V^2}{R}$$

Examples

1. In the circuit of fig. 11 find the power expended in the 5.1-ohm resistance.

As the current through the resistance is known, use the relation $P = I^2 R$ giving

$$P = (0.5)^2 \times 5.1 = \frac{5.1}{4} = 1.275 \text{ watts.}$$

2. At what rate is chemical energy being converted into electrical energy in the battery?

The energy transformed in T seconds is EIT joules, hence the rate of conversion is EI joules per second or EI watts. This is equal to 10×0.5 or 5 watts.

3. In what time would 1 B.O.T.U. of electrical energy be expended in the whole circuit ?

$$1 \text{ B.O.T.U.} = 1,000 \text{ watt-hours}$$

Energy is being expended at a rate of 5 watts, hence the time taken to expend 1,000 watt-hours would be $\frac{1,000}{5} = 200$ hours.

4. A certain motor gives an output of 50 horse power. Assuming it has no losses, find the annual cost of running 8 hours a day 300 days per annum, if electrical energy costs 1d. per B.O.T.U.

$$50 \text{ H.P.} = 50 \times 746 \text{ watts}$$

$$\text{Total energy required} = 50 \times 746 \text{ watts} \times 300 \times 8 \text{ hours}$$

$$= 50 \times 2,400 \times 746 \text{ watt-hours}$$

$$= 5 \times 24 \times 746 \text{ Kilowatt-hours or B.O.T.U.}$$

$$\text{Cost of energy at 1d. per B.O.T.U.} = 5 \times 24 \times 746 \text{ pence}$$

$$= 7,460 \text{ shillings}$$

$$= \underline{\underline{£373.}}$$

40. (i) If no mechanical work is performed the whole of the electrical energy supplied to a circuit is expended in producing heat. The only exception to this rule is when electromagnetic radiation is produced, e.g. light or wireless waves. The relation between the energy expended and the heat produced was discovered by the British scientist Joule, who performed a very large number of experiments in which a quantity of water was stirred continuously by various mechanical methods, and the increase in the temperature of the water was measured. He found that about 780 ft.-lb. of work is required to raise the temperature of 1 lb. of water through 1° Fahrenheit. In the C.G.S. system the unit of heat is the calorie, which is the quantity of heat necessary to raise the temperature of one gram of water by 1° Centigrade. The calorie is equal to 4.2 joules, and as the heat developed by a current of I amperes flowing through a resistance of R ohms for

t seconds is $I^2 R t$ joules, it is also equal to $\frac{I^2 R t}{4.2}$ calories.

(ii) When the temperature of a body is increased by $\theta^\circ \text{C.}$, the amount of heat gained (in calories) is equal to product of the mass of the body in grams, the increase of temperature, and to a constant known as the specific heat of the material of which the body is composed. This constant is defined as the ratio

$$\frac{\text{Heat required to raise 1 gram of the substance through } 1^\circ}{\text{Heat required to raise 1 gram of water through } 1^\circ}$$

The specific heat of water is obviously unity, and is much less than unity for all metals. If m is the mass of a body and s its specific heat, an increase in its temperature of $\theta^\circ \text{C.}$ causes the body to gain a quantity of heat, h , and

$$h = m s \theta \text{ calories.}$$

This equation, together with the relation 1 calorie = 4.2 joules, enables the electrical engineer to make calculations regarding the cost of heating by electrical means, as in the following example.

If electrical energy is 2d. per B.O.T.U. find the cost of boiling 10 gallons of water, if its initial temperature is 25°C. , assuming no heat is wasted.

CHAPTER I.—PARA. 41

Temperature rise = 75°C .

The mass of 1 gallon of water is 10 lb., and of 10 gallons, 100 lb. or 100×453.6 grams.

$$\begin{aligned} h &= m \theta \text{ since for water } s = 1 \\ \therefore h &= 453.6 \times 100 \times 75 \text{ calories} \\ &= 4.2 \times 453.6 \times 100 \times 75 \text{ joules} \\ &= \frac{4.2 \times 453.6 \times 100 \times 75}{3,600 \times 1,000} \text{ Kilowatt-hours or B.O.T.U.} \\ &= 4 \text{ B.O.T.U. approx., costing eightpence.} \end{aligned}$$

In many cases the conversion data given in Table VII, Appendix A, will be found of assistance. The above example may be worked as under.

From the Table it is found that one H.P.-hour is the amount of energy required to raise 17.2 lb. of water from 62°F . to 212°F ., that is through 150°F .

As $25^{\circ}\text{C} = 77^{\circ}\text{F}$., the water must be raised only $212^{\circ} - 77^{\circ} = 135^{\circ}\text{F}$. 1 H.P.-hour will raise $\frac{150}{135} \times 17.2 = 19.1$ lb. of water through this range of temperature.

To raise the temperature of 100 lb. will require $\frac{100}{19.1}$ H.P.-hour.

$$\begin{aligned} &= \frac{746 \times 100}{19.1} \text{ watt-hour.} \\ &= 3.9 \text{ B.O.T.U. at a cost of } 7.8 \text{ pence.} \end{aligned}$$

Electric batteries

41. An electric battery consists of any number of primary or secondary cells arranged either in series, in parallel, or in a series parallel combination. Each cell may be considered to have an E.M.F. of E volts and an internal resistance of r ohms. If a number s of similar cells are arranged in series, the total E.M.F. of the combination will be sE volts and the total internal resistance sr ohms. If this battery is connected to an external circuit having a resistance of R ohms, the current in the circuit will be

$$I = \frac{sE}{R + sr} \text{ amperes.}$$

On the other hand if a number p of similar cells are connected in parallel the E.M.F. is not increased but remains equal to that of a single cell, in fact the sum of all the positive elements may be considered to form a single positive element and the negative elements may be treated similarly. The effect of the parallel group is therefore that of a single cell with elements p times as large as those of one cell, and although the E.M.F. is not increased the internal resistance is decreased, being only $\frac{1}{p}$ th of the resistance of each cell. The current in the external circuit is

$$I = \frac{E}{R + \frac{r}{p}},$$

a current $\frac{I}{p}$ flowing through each cell.

If p groups, each consisting of s cells in series, are arranged in parallel, each group of cells in series may be considered to form a single cell of E.M.F. sE volts, and internal resistance sr ohms. A number p these groups in parallel will then have E.M.F. sE volts and internal resistance $\frac{sr}{p}$ ohms. If connected to an external resistance of R ohms the current in the external circuit is

$$I = \frac{sE}{R + \frac{sr}{p}} \text{ amperes.}$$

If only a certain number of cells are available, for a given resistance in the external circuit there is always some method of arranging the cells which will give the greatest value of current, and consequently the maximum expenditure of power in the external circuit. It is found that the best arrangement for the fulfilment of these requirements is such that $\frac{rs}{p} = R$, that is, the total effective internal resistance should be equal to the external resistance if maximum current is required. This condition will also make I^2R a maximum, and therefore is the best arrangement when it is desired to dissipate the greatest possible power in the external circuit. If the number of cells available is n , and they are arranged s in series and p in parallel, $n = sp$, or $\frac{s}{p} = \frac{n}{p^2}$.

For maximum heating effect

$$\begin{aligned} R &= \frac{s}{p} r \\ R &= \frac{n}{p^2} r \\ p^2 &= \frac{nr}{R} \\ \therefore p &= \sqrt{\frac{nr}{R}} \end{aligned}$$

Example.—20 cells each of internal resistance $\cdot 1$ ohm are to be arranged to give maximum current through a resistance of $\cdot 08$ ohms. What is the best arrangement?

The best arrangement is p groups in parallel of s cells in series, where

$$\begin{aligned} p^2 &= \frac{20 \times \cdot 1}{\cdot 08} \\ &= \frac{2}{\cdot 08} \\ &= \frac{200}{8} = \frac{100}{4} \\ p &= \frac{10}{2} = 5 \end{aligned}$$

Answer.—5 parallel groups of 4 in series.

If the E.M.F. of each cell is 2 volts, the current with this arrangement will be

$$\frac{sE}{2R} = \frac{8}{\cdot 16} \text{ or } 50 \text{ amperes.}$$

CONDUCTORS AND RESISTANCES

42. Conductors are used for two different purposes, (i) to convey electrical energy to the point at which it is to be utilised, and (ii) to control and regulate electrical currents and voltages. In the former application the resistance of the conductor is a disadvantage as it entails a loss of energy. This loss of energy takes place at a rate of I^2R joules per second, I being the mean value of the current and R the resistance of the conductor. For this reason, only materials of low specific resistance are used for "supply leads", as such conductors are generally termed, copper being employed for preference, although considerations of tensile strength sometimes necessitate the use of iron wire and considerations of weight sometimes cause aluminium to be employed, e.g. in long spans of overhead power lines. In the second application of conductors, the control is obtained by variation in the amount of resistance included in the circuit, and high specific resistance is then a desirable property of the conductor, as it allows a smaller and cheaper design than would be possible with a good conductor. Resistances used for control purposes

CHAPTER I.—PARA. 43

receive many special names, but can generally be classed as rheostats, potentiometers and fixed resistors, while two broad divisions may also be made according to the character of the conductor employed, thus giving rise to "wire wound" and "composition" types of resistances. Wire wound resistances are almost invariably manufactured from one of the high resistance alloys. These are substances of high specific resistance and almost negligible temperature coefficient, the principal being platinum silver, platinoid, German silver, manganin and eureka. The specific resistance of these alloys is from fifteen to thirty times that of copper, while the temperature coefficient is less than one-tenth that of copper. The advantage of the negligible temperature coefficient is that the resistance of the circuit is not appreciably dependent upon the amount of current flowing, as is the case when a substance of high temperature coefficient is used. Particulars of the alloys mentioned above will be found in Table I, Appendix A.

Rheostats and potentiometers

43. (i) A variable resistance arranged in series with other devices in order to effect a control of current flowing through the latter, is called a rheostat. For currents up to about 20 amperes, the resistance unit generally takes the form of a slab or tube of insulating material, upon which is carried a winding of eureka or manganin wire. A sliding contact is arranged to make electrical contact with the wire at any point along the length of the slab or tube, or else a number ofappings is brought from the wire at any point along the length of the slab or tube, or else a number ofappings is brought from the wire to a row of contacts, connection to the latter being effected by a rotating arm. These are shewn diagrammatically in fig. 17.

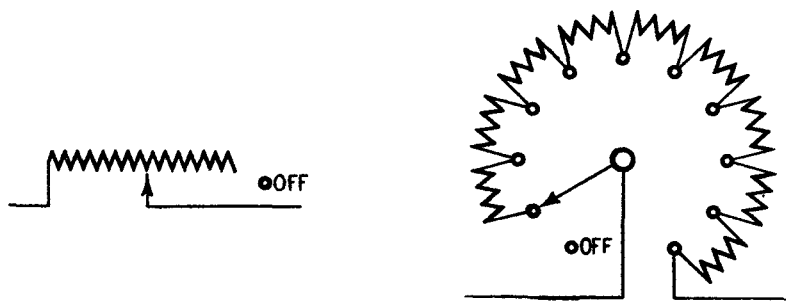


FIG. 17, CHAP. I.—Rheostats.

(ii) A potentiometer is a variable resistance arranged in such a manner that a certain proportion of the available P.D. may be applied to a particular instrument. In fig. 18a is shewn a 2-volt accumulator, and a potentiometer so connected as to allow any fraction of the available E.M.F. to be applied to the instrument X, the nature and purpose of the latter being immaterial

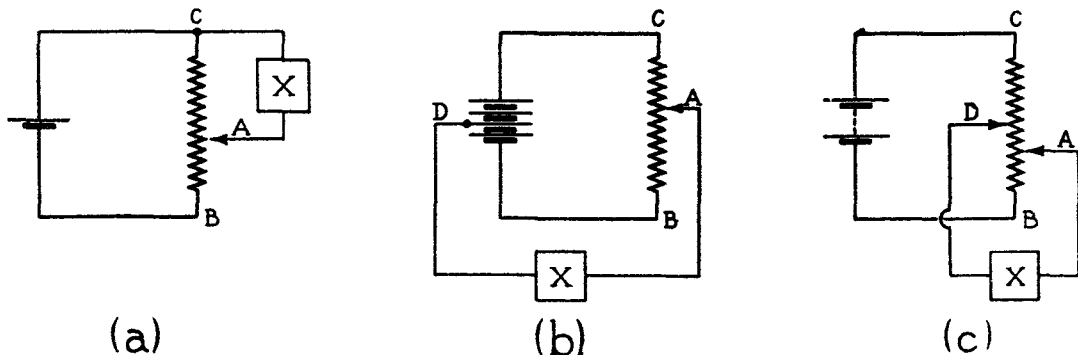


FIG. 18, CHAP. I.—Potentiometers.

COLOUR CODE FOR SMALL RESISTANCES USED IN RADIO APPARATUS

Body	Tip	Spot (or band)
0	0	—
1	1	0
2	2	00
3	3	000
4	4	0,000
5	5	00,000
6	6	000,000
7	7	0,000,000
8	8	00,000,000
9	9	000,000,000

Example



Explanation

The colours used have the numerical values indicated above, the numbers being set down in the order Body, Tip, Spot. In the example, the body, being green, gives the number 5. The tip (red) indicates 2 and the spot (blue) shows that six noughts follow. The value of the resistance in ohms is therefore 52,000,000

Fig. 19.

to the present discussion. When the slider A is carried to the end of the winding B, the whole of the available 2 volts is applied to X, but on carrying the slider to position C, no voltage is applied to X, while at any intermediate point the voltage applied to the instrument will lie between these limiting values. Two methods of connecting a potentiometer by which a reversal of direction of applied voltage can be obtained, are shewn in fig. 18b and c. The first necessitates a battery of twice the maximum voltage which it is desired to apply to the instrument. The second method overcomes this disadvantage.

44. Resistors, or fixed resistances, are often fitted in radio instruments for different purposes. Many of these are of the order of 10^5 or 10^6 ohms, and it is necessary to obtain this value in as small a space as possible. In some instances a large number of turns of very fine wire (having an insulating surface formed by oxidisation) are wound in spiral form upon a cord about $\frac{1}{16}$ in. in diameter, the ends of the wire being pressed into metal connectors and the cord covered by a tube of insulating material. In another type an extremely fine wire is wound in sections upon a fireclay tube, being afterwards covered with porcelain and "fired" in order to glaze the covering. These are called vitreous resistances, although the term vitreous does not apply to the resistance element but to the glass-like exterior. Such resistances are made in units of from a few ohms to one hundred thousand ohms. For resistances of a higher order but very low current-carrying capacity, a metallic film of minute thickness may be applied to a rod of insulating material, connecting leads being connected to the ends and the whole enclosed in a glass or cardboard tube. Suitable steps are taken to exclude moisture, e.g. by coating with paraffin wax or similar material. When resistances of this design are made, a colour code is adopted to denote the value of the resistance. The resistance carried three colours placed in the positions indicated on the typical resistance illustrated in fig. 19. Ten different colours are used, and are read in the order "body", "tip", and "spot" or band. The numerical values and method of reading the code are given in the figure. The use of this code is practically confined to high value resistors of small physical dimensions which form component parts of modern radio apparatus.

Fuses

45. Fuses are protective devices, which are generally placed in series with circuits in such a manner as to interrupt the circuit if the current exceeds a certain value. The fuse consists of a short length of wire, generally of tin or an alloy of tin and lead. If the current becomes excessive, the heat generated in the fuse will be sufficient to melt it, and the circuit is broken. In order that the hot metal shall not spray out and ignite any inflammable material in the vicinity, it is usual to enclose the wire in a porcelain holder.

THE CARBON MICROPHONE

46. This instrument is dealt with in this chapter because it is essentially a special form of resistance, the value of which can be controlled by sound waves. In its simplest form it consists of two polished carbon plates which serve as terminal electrodes, and between them is placed a small quantity of granulated carbon. One of the electrodes may be rigidly fixed to some support, and the other electrode is then mounted upon a flexible insulating diaphragm, mica being usually employed. Fig. 20 shews the carbon microphone in section, and also the manner in which it is connected to a battery and telephone receiver in order to form a simple telephone. The action of the instrument is as follows:—When a sound wave impinges upon the flexible diaphragm, the alternate compression and rarefaction of the atmosphere causes corresponding movement of the diaphragm, and the resistance of the microphone varies accordingly. In the absence of a sound wave, the microphone carries a steady current from a suitable battery, which is also caused to flow through the windings of a telephone receiver. Variation of the resistance of the microphone causes corresponding variations of the current flowing through the telephone receiver and the diaphragm of the latter is set into vibration by the variations of current through the winding, setting up in the surrounding air a sound wave having characteristics similar to

CHAPTER I.—PARA. 47

the original sound. The action of the telephone receiver is dealt with in Chapter II, and a more complete description of a service type of microphone designed for use in aircraft is given in A.P. 1186 (Signal Manual, Part IV, Section V, Chapter XII).

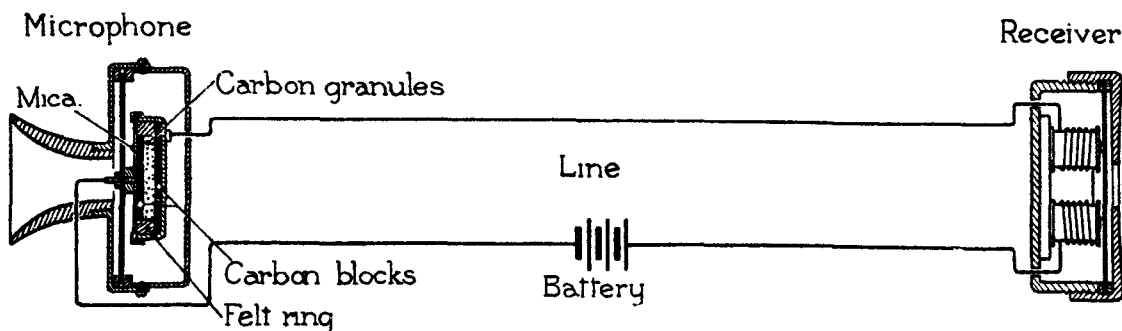


FIG. 20, CHAP. I.—Simple telephone.

ELECTROSTATICS

47. A charged body has already been defined as one which has gained or lost electrons. The forces of attraction and repulsion exhibited by such bodies are exerted in a region surrounding them, which is called a field of electric force or an electric field. It may be imagined to consist of an infinite number of lines of electric force, a line of force being a line which shews the direction in which the force is acting at all points along its length. A field of force is shewn diagrammatically in fig. 21, the particular case chosen being that of a charged sphere remote from all other bodies.

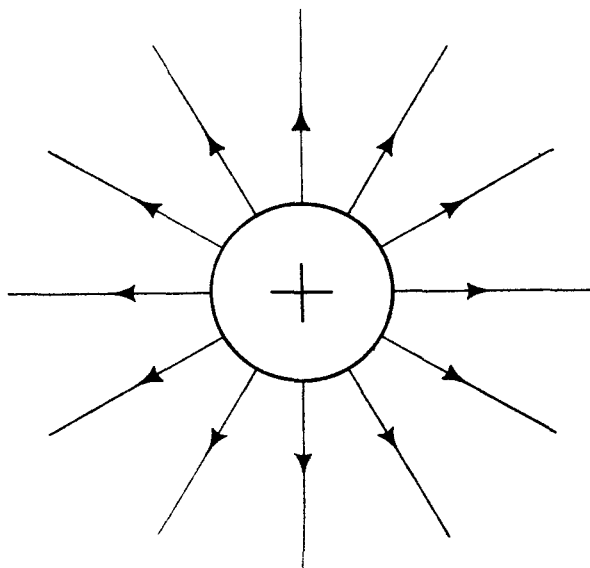


FIG. 21, CHAP. I.—Field of electric force.

The size and nature of the charge are denoted by a number denoting the number of unit charges carried and a + or - sign to denote positive or negative charge respectively. A further convention is the arrow head shewn on a number of the lines, in order to indicate the direction in which a positive charge would be urged if placed in the field.

Electric flux and flux density

48. The electric field may be considered as an electric stress applied to the medium surrounding the charge, which in its turn causes the medium to undergo a strain or displacement. This strain, when speaking of the whole area of the field perpendicular to the direction of the force, is called the electric flux. The word "stress" has a definite technical meaning although it is often misused. In mechanics, stress is any force (per unit area) applied to a body in such a manner as to alter its shape or size, the resulting fractional change in dimensions of the body being called the strain. Under certain conditions, the strain produced (D units) is proportional to the stress applied (F units) and the numerical relation between stress and strain may then be expressed by the equation $F = KD$, where K is a constant for any particular material, and is called the modulus of elasticity. Now the stress F and the strain D are of an entirely different physical nature, and in like manner, it is considered that the electric stress E , also called the electric field strength*, causes an electric strain or displacement D , which is also called the electric flux density. If any small area is taken in the electric field, perpendicular to its direction, the lines of force through all adjacent points bounding this area may be considered to enclose a tube of electric flux. These tubes of flux may be thought of as elastic bands, tending to contract in the direction of their length while expanding in their cross-section, this tendency being due to the applied force. If one end of a tube of flux is situated upon a positive charge, the opposite end must be terminated upon an equal and opposite charge. The imaginary elastic property of the tube then tends to draw the two charges together, and accounts for the attraction between like charges. Repulsion may also be explained by the tendency of the tubes to increase in cross-section. Two adjacent tubes of unlike sign tend to unite, contracting in such a way as to form a single tube. This conception is illustrated in fig. 22, which shews various stages in the contraction of the field between two unlike charges, as they approach each other.

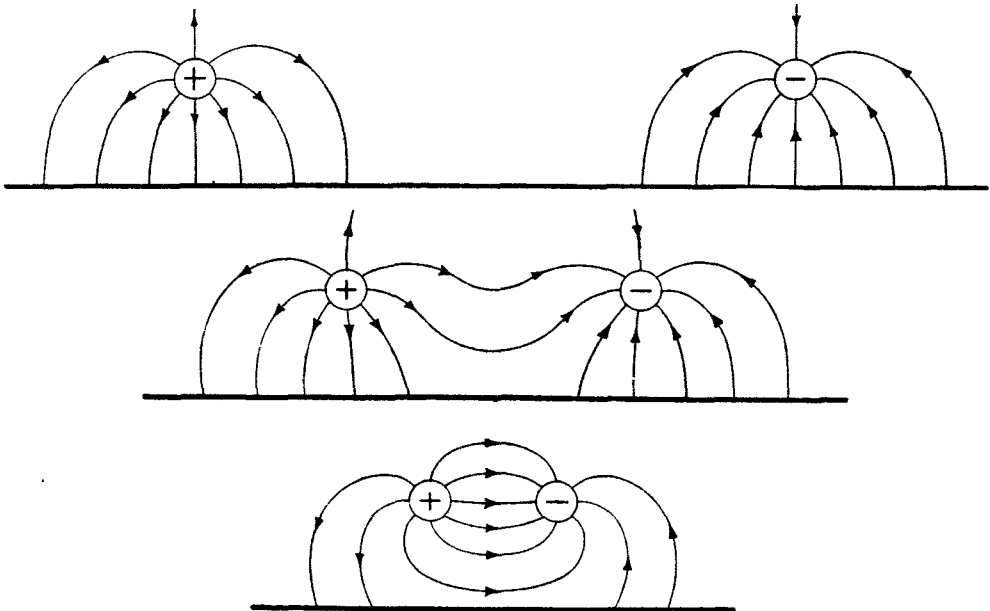


FIG. 22, CHAP. I.—Fields between unlike charges.

Electrostatic system of units

49. It has been found experimentally that the force exerted between two charged bodies varies inversely as the square of the distance between them. The system of units based on these forces is known as the electrostatic system (E.S.U.).

* The symbol ϵ (epsilon) is sometimes used to denote electric field strength, but in this publication E (gamma) has been adopted in order to avoid confusion with the base of Napierian logarithms.

CHAPTER I.—PARA. 50

The basis of this system is the electrostatic unit of quantity, or electrostatic unit charge, which is defined as follows. If two equal charges placed one centimetre apart in vacuo exert upon each other a force of one dyne, then each is a unit charge. The E.S.U. of quantity is a very small unit, being $\frac{1}{3 \times 10^9}$ of the practical unit or coulomb. A single unit charge can be isolated as follows. A pith ball, exactly 2 cm. in diameter, and covered with gold leaf in order to make its surface conductive, is suspended in the middle of a large room by means of a single silk fibre. A battery having an E.M.F. of 300 volts has its positive terminal connected to the wall of the room, and the pith ball is then touched with the end of a fine wire which is connected to the negative terminal of the battery. The pith ball then acquires a charge of one E.S.U. Two such charges placed one cm. apart in vacuo repel each other with a force of one dyne. If larger charges are given, say Q_1 and Q_2 E.S.U. respectively, then under the same conditions the force exerted is $Q_1 \times Q_2$ dynes, and in general, the force exerted between two charges Q_1 , Q_2 , the distance apart being d cms. (in vacuo) is $\frac{Q_1 \times Q_2}{d^2}$ dynes.

If the charges are situated in some material medium, the force exerted is not the same as if they are situated in vacuo. For instance, if they are placed in pure vaseline—which is almost a perfect insulant and therefore will allow practically none of the charge to leak away—the force would be reduced to about one half that exerted in vacuo. On the other hand, in air the forces are practically the same as in vacuo. The force exerted between charged bodies thus depends upon a property of the medium in which they are situated. This property is called its “dielectric constant” or “permittivity” or formerly its “specific inductive capacity”. The symbol for permittivity is κ .

The complete relation between the force and the charges in any medium of permittivity κ is therefore

$$\text{Force} = \frac{Q_1 Q_2}{\kappa d^2} \text{ dynes.}$$

Electric field strength

50. The strength of an electric field at any point is defined as the force exerted upon a unit charge placed at that point. The electric field strength at a distance d from a point charge of Q units in a medium of permittivity κ is

$$F = \frac{Q}{\kappa d^2} \text{ dynes per unit charge.}$$

One unit tube of electric flux is assumed to start from a unit positive charge, or to end on a unit negative charge. The total number of unit tubes of electric flux emanating from a charge is the electrostatic flux or electric flux. (Symbol Ψ).

The electrostatic flux density or electric flux density is the amount of electrostatic flux per unit area normal to the direction of the flux. Its symbol is D .

$$\text{Thus } D = \frac{\Psi}{A}$$

where Ψ is the flux (assumed uniform) through an area A square centimetres.

Since one tube of electric force is associated with each unit charge, the total flux over a surface enclosing a charge of Q units must be Q tubes. If then a sphere of radius r centimetres encloses a point charge of Q units, the surface area of the sphere being $4\pi r^2$ square centimetres, the electrostatic flux density at this surface is

$$D = \frac{Q}{4\pi r^2} \text{ tubes per square centimetre}$$

but the electric field strength F is equal to $\frac{Q}{\kappa r^2}$,

$$\text{so that } F = \frac{4 \pi r^2 D}{\kappa r^2} = \frac{4 \pi D}{\kappa}$$

$$\text{or } D = \frac{\kappa F}{4 \pi} \text{ tubes per square centimetre.}$$

The dielectric constant or permittivity

51. The permittivity of a perfect vacuum is assumed to be unity. For all insulating substances or dielectrics, it is greater than unity, and does not vary appreciably at ordinary temperatures with slight temperature changes. Dry air has a permittivity very slightly greater than unity. Table II, Appendix A gives the value of this constant for several common substances. The reader is warned, however, that the permittivity of a material does depend upon the amount of ionisation or upon the number of free electrons present. Thus certain regions of space between the earth and the sun appear to have dielectric constants less than unity. Again, suppose two bodies carrying unlike charge were placed in a conducting medium, then the surplus electrons composing the one charge would flow to the body having a deficit of electrons, and therefore after a very short—in fact infinitesimal—period there would be no force between them. From this aspect, the permittivity of a perfect conductor is infinitely great. The principal occasion upon which this effect is of importance is in the consideration of the travel of wireless waves in the upper regions of the atmosphere, which is dealt with later.

Energy stored in an electric field

52. If one of the charged bodies hitherto considered were fixed, while the other were free to move it is obvious that under the influence of the electric force, motion would take place, or work would be done. Hence the electric field possesses the capability to do work—which is our conception of energy. Whenever an electric field is established, potential energy is stored. If motion does take place, this potential energy is converted into kinetic energy; and current flows from a point of higher potential to a point of lower potential. This leads to the notion of a difference of potential between the two points, just as in the case of conduction current. The P.D. between two points in an electric field is the work done when a unit charge moves from one point to the other. If the movement of a unit positive charge is due to the electric field, then the first point is at a higher potential than the second. If the unit positive charge is moved against the opposition of the electric field, then the first point is at a lower potential than the second. The earth's surface is assumed to be at zero potential, and the potential of any point in the field can be stated with reference to this surface. The E.S.U. of P.D. is the P.D. between two points if one erg of work is performed in moving a unit charge from one point to the other, thus

$$\text{E.S. Unit P.D.} = \frac{1 \text{ erg}}{\text{E.S. unit charge}}$$

The practical unit of P.D. is the volt, and the practical unit of work the joule ($= 10^7$ ergs).

$$\begin{aligned} \text{Hence one E.S. unit of P.D.} &= \frac{\frac{1}{10^7} \text{ joule}}{\frac{1}{3 \times 10^9} \text{ coulomb}} = 300 \text{ joules per coulomb.} \\ &= \underline{\underline{300 \text{ volts}}} \end{aligned}$$

The significance of the 300-volt battery used to obtain an E.S. unit charge will now be appreciated.

CHAPTER I.—PARAS. 53-55

Capacitance

53. The potential of a charged body is proportional to its charge. This relation may be written $Q \propto V$ or $Q = CV$, where C is a constant. This constant, the ratio of charge to potential, is called the capacitance of the body. The charge which a given body can hold at a given potential can be increased by concentrating the region in which its field exists. An arrangement by which the capacitance is thus increased is called a condenser. In effect, a condenser may be considered as an arrangement of conductors in which the tubes of flux are packed more closely together, and the extra work required to accomplish this appears as an increase of potential energy. The capacitance of some bodies is easily determined. Thus consider a sphere of radius r , remote from all other bodies in space. If its surface is given a charge of Q units, the work which would have to be done in order to concentrate that charge on its centre is the potential at the centre. That is

$$V = \frac{Q}{r^2} \times r = \frac{Q}{r}$$

because $\frac{Q}{r^2}$ is the force at the centre and r is the distance, while work = force $\times$ distance.

Since Capacitance is the ratio $\frac{\text{Charge}}{\text{Potential}}$, or $\frac{Q}{V}$, we have $C = \frac{Q}{V} = Q \div \frac{Q}{r} = r$, hence the capacitance of a sphere is equal to its radius; the E.S. unit of capacitance is also called the centimetre. A body of unit capacitance has a potential of one E.S. unit when it is given unit charge. In isolating our unit charge, a pith ball (assumed spherical and of radius 1 cm.) was given a potential (with respect to earth or zero potential) of 300 volts or 1 E.S.U. Hence it acquired a charge of 1 E.S.U. The E.S. unit of capacitance is inconveniently small for most commercial purposes, and the practical unit of capacitance is the farad; a body has a capacitance of one farad if a charge of one coulomb raises its potential by one volt. However, this unit is too large for actual use and the microfarad or one of its subdivisions is generally employed.

$$1 \text{ coulomb} = 3 \times 10^9 \text{ E.S.U. of quantity}$$

$$1 \text{ volt} = \frac{1}{3 \times 10^9} \text{ E.S.U. of potential}$$

Hence

$$1 \text{ farad} = 9 \times 10^{11} \text{ E.S.U. of capacitance}$$

$$1 \text{ microfarad} = \frac{1}{10^6} \text{ farad}$$

(μF)

$$1 \text{ millimicrofarad} = \frac{1}{10^9} \text{ microfarad}$$

(m. μF)

$$1 \text{ micro-microfarad} = \frac{1}{10^{12}} \text{ microfarad}$$

($\mu\mu F$)

$$= \frac{1}{10^{12}} \text{ farad}$$

Capacitance of a condenser

54. A common form of condenser consists of two parallel plates separated by, or immersed in, an insulating material. If one of these plates is given a positive charge, and the other a negative charge, the plates themselves being remote from each other, the lines of flux from each plate would terminate on the earth. If the two plates are now brought near to each other, the lines of force can shorten and link from one plate to the other, so that the electric field is considerably concentrated.

55. Consider an uncharged parallel plate condenser connected to a battery, switch and galvanometer, the latter being an instrument which indicates the presence and direction of a current.

First, let the plates be well separated as in fig. 23a. On closing the switch, the electromotive force of the battery will urge electrons along the circuit, so that electrons flow from the plate A through the battery and to plate B. This current of electrons will be shown by the deflection of the galvanometer. When the P.D. between the plates is equal to the E.M.F. of the battery, there will be no further flow of electrons. This will occur a fraction of a second after closing the switch. The two plates are now oppositely charged, plate A having a deficit of electrons, and plate B a surplus. At the same time an electric field is established between the plates in the medium separating them, and a displacement of electrons will occur in this medium, the electrons trying to move toward the plate A, which has a deficit of electrons. As has already been stated, although the electrons in an insulating material cannot leave their own atoms, they are strained in their orbits, and this slight movement constitutes a displacement current.

If now the two plates are brought more closely together as in fig. 23b, the negative charge on plate B tends to nullify the positive charge on plate A so that the P.D. between the plates is decreased and a momentary electron flow takes place into plate B, restoring the P.D. to equality with the E.M.F. of the battery. On breaking the switch, each plate is left in a charged state, the P.D. between the plates being equal to that of the battery from which the condenser was charged.

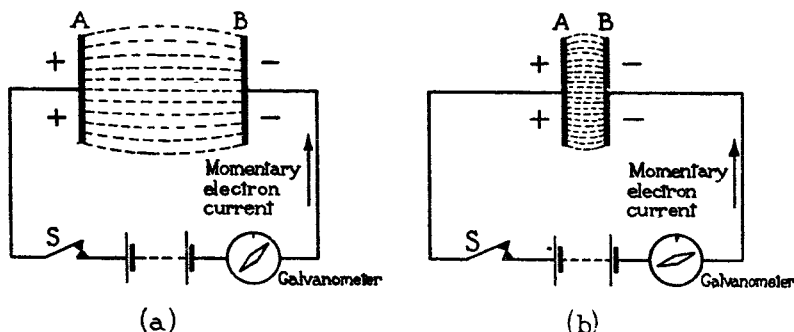


FIG. 23, CHAP. I.—Increase of capacitance of condenser.

If now the plates are connected by a conductor, the surplus of electrons on plate B will surge into the conductor displacing other electrons and causing a conduction current to flow. Electrons from the end of the conductor nearest to plate A will flow into the plate and unite with the atoms which are deficient in electrons, restoring them to their neutral state. At the same time, the strain on the electrons in the material of the dielectric will be released, the electrons will recover their normal orbits, and in so moving constitute a displacement current in the reverse direction to the displacement current set up by the charging process.

Capacitance of a parallel plate condenser

56. The capacitance of a parallel plate condenser may be derived as follows :—

Assuming that the charge on each plate is equally distributed over its surface, the lines of force will be parallel and the flux density uniform.

Let V = the P.D. between the plates (E.S.U.)

Q = the charge on one plate (E.S.U.)

d = the distance between plates in cm.

A = the area of plates in cm.².

Then the flux density $D = \frac{Q}{A} \frac{\text{tubes}}{\text{cm.}^2}$.

The electric field strength is E , and

$$E = \frac{4\pi D}{\epsilon}$$

CHAPTER I.—PARAS. 57-58

The P.D. between the plates = the work done in moving a unit charge against a force F through a distance d .

$$\begin{aligned}\text{i.e. } V &= Fd \\ \therefore V &= \frac{4\pi D}{\kappa} \times d \\ &= \frac{4\pi Q}{\kappa A} \times d\end{aligned}$$

There is a further relationship between V and Q , namely

$$V = \frac{Q}{C} \text{ or } C = \frac{Q}{V}$$

Hence
$$C = \frac{Q}{\frac{4\pi Q d}{\kappa A}} = \frac{A \kappa}{4\pi d} \text{ (E.S.U.)}$$

If C is to be expressed in microfarads,

$$C = \frac{A \kappa}{4\pi d \times 9 \times 10^5} \text{ microfarads.}$$

Energy stored in a condenser

57. It has already been stated that work must be done in order to establish an electric field and this work is stored in the field as potential energy. The amount of energy stored in a condenser of C farads when charged to a P.D. of V volts will now be calculated. The charge introduced into the condenser is given by the equation $Q = CV$. The average charging current, during the time t taken to charge the condenser will be $\frac{Q}{t}$ or $\frac{CV}{t}$ and since the P.D. at beginning

of charge is zero, while at the end it is V , the average voltage during charge is $\frac{V}{2}$. The average rate of doing work is therefore $\frac{CV}{t} \times \frac{V}{2}$ joules per second, and the total work done in t seconds is therefore $\frac{CV^2}{2}$ joules. This work done is stored as energy in the charged condenser.

A useful analogy is the storage of air in a closed metal cylinder by means of an air pump. Let the cylinder have a capacity of C cubic feet at atmospheric pressure, then it will hold $C \times V$ cubic feet at a pressure V . Now, to force the first additional particle of air into the cylinder requires practically no pressure. When the cylinder has been pumped up to a pressure of V lb. per square foot, a pressure of just over V Lb. per square foot will be necessary in order to force in another particle of air. Thus the average pressure during the whole operation will be $\frac{V}{2}$ Lb. per square foot and the total work done will be

$$\frac{V \text{ Lb.}}{2 \text{ ft.}^2} \times CV \text{ ft.}^3 = \frac{CV^2}{2} \text{ ft. Lb.}$$

and this is the amount of work which could be done by the compressed air.

Energy density

58. It is sometimes convenient to consider the amount of energy stored in a dielectric without reference to the capacitance. This may be illustrated by a consideration of the parallel plate condenser, the field being assumed to be parallel and uniform.

The capacitance of such a condenser in E.S.U. is $\frac{A}{4\pi d}$ units and the energy stored is $\frac{1}{2} CV^2$ ergs, if V is the P.D. between the plates, also in E.S.U.

Hence the energy stored $= \frac{1}{2} V^2 \times \frac{A \kappa}{4 \pi d}$ ergs and since the total volume of the dielectric is Ad , this is equivalent to an energy density of $\left(\frac{V^2}{2} \times \frac{A \kappa}{4 \pi d} \right) \div Ad$ ergs per cubic centimetre.

$$\text{Now} \quad \left(\frac{V^2}{2} \times \frac{A \kappa}{4 \pi d} \right) \div Ad = \frac{\kappa}{8 \pi} \left(\frac{V}{d} \right)^2$$

V is the work done on a unit charge in moving it from one plate to the other.

$\frac{V}{d}$ is the work done per centimetre, i.e. is the electric field strength F of the field.

Hence the energy density in a uniform field of strength F is $\frac{\kappa F^2}{8 \pi}$ ergs per cubic centimetre.

The conception of electric field strength as the P.D. per centimetre leads to a practical unit of electric field strength—the volt per centimetre.

One E.S.U. of field strength = 300 volts per centimetre. A submultiple of this, the millivolt per metre, is frequently used to measure the strength of an electric field due to a distant radio transmitter, at any point where wireless reception is contemplated.

Charge and discharge of a condenser

59. It is now proposed to consider the phenomena associated with the charging of a condenser through a resistance. The circuit shewn in fig. 24 comprises a source of constant E.M.F. of E volts, a condenser of C farads and a resistance of R ohms. The condenser and resistance (in

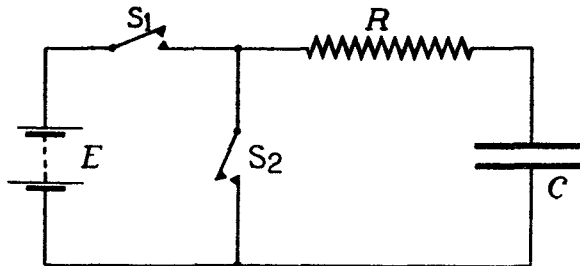


FIG. 24, CHAP. I.—Condenser and resistance in series.

series) can be connected to the source of supply by means of the switch S_1 . Before this switch is closed the condenser possesses no charge and the P.D. between its plates is zero. On closing the switch a current will commence to flow, charging the condenser, the magnitude of this current at the instant of completing the circuit being $\frac{E}{R}$ amperes: note particularly that as the condenser offers no "back pressure" or counter-E.M.F., the initial charging current is the same as if the condenser were short circuited. In a very short interval of time, however, the condenser will receive a charge, e.g. in t seconds a charge q , equal to $I \times t$ or $\frac{E}{R} t$ coulombs, and will therefore exert a counter-E.M.F., e , equal to $\frac{E}{CR} t$ volts, because $e = \frac{q}{C}$ and $q = \frac{E}{R} t$. In order to shew the effect, let us assume that $E = 200$ volts, $C = 10 \mu F$, $R = 1,000$ ohms. Then when $\cdot 0025$ second has elapsed from the time of closing the switch the counter-E.M.F. will be $\frac{200}{1,000} \times \frac{10^6}{10} \times \cdot 0025$ which is 50 volts. Now the final charge of the condenser will be CE or $\frac{1}{10^6} \times 200 = \cdot 002$ coulombs, and if the initial rate of charge were maintained the voltage of the condenser would reach its final value E in a time T , where $\frac{E}{CR} T = E$ or $T = CR$ seconds.

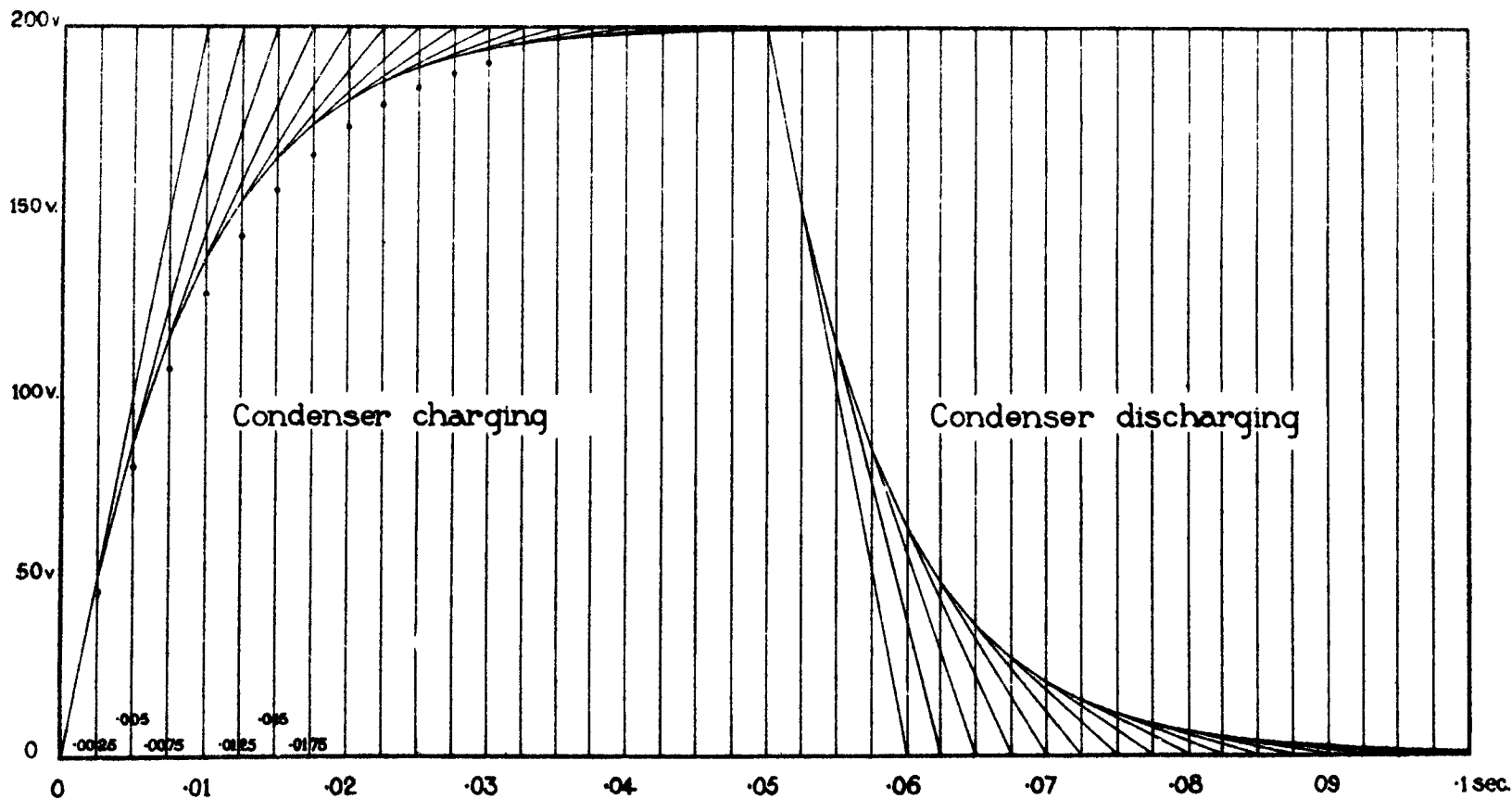
CHAPTER I.—PARA. 59

The time which would be taken to charge the condenser to a voltage equal to that applied, if the initial rate of charging could be maintained, is called the time constant of the circuit. In practice the condenser does not charge at the initial rate, because the counter-E.M.F. opposes the applied voltage, and therefore the charging current falls as the counter-E.M.F. increases. Thus after an interval $\cdot 0025$ second, the counter-E.M.F. is 50 volts, and only 150 volts are available to cause a further charge. During a second interval of time of the same duration, the current which will flow will be $\frac{150}{1,000}$ amperes, the charge $\cdot 15 \times \cdot 0025$ coulombs, and the increase

of voltage $\frac{Q}{C}$ or $\cdot 15 \times \cdot 0025 \times 10^5$ which is 37·5 volts. At the end of the second interval the counter-E.M.F. will therefore be $50 + 37\cdot 5$ or 87·5 volts. The voltage available to cause a further charge is now reduced to 112·5 volts, and at the end of yet another interval of $\cdot 0025$ second the current will be $\cdot 1125$ ampere, the additional charge $\cdot 1125 \times \cdot 0025$ coulomb, and the increase of counter-E.M.F. $\cdot 1125 \times \cdot 0025 \times 10^5$ or 28·125 volts. The total counter-E.M.F. at the end of the third interval is therefore $87\cdot 5 + 28\cdot 125 = 115\cdot 625$ volts. Proceeding in this manner we may complete a table as follows, and plot the results on squared paper. The resulting curve will then shew the rate at which the condenser P.D. rises.

Interval	Average Current, $i = \frac{E - e}{R}$	Charge in coulombs during interval $q = i t$	Counter-E.M.F. produced during interval $\frac{q}{C}$	Total counter-E.M.F.	Voltage available for further charge
1	$\cdot 2$	$\cdot 0005$	50	50	150
2	$\cdot 15$	$\cdot 000375$	37·5	87·5	112·5
3	$\cdot 1125$	$\cdot 00028125$	28·125	115·625	84·375
4	$\cdot 084375$	$\cdot 000212$	21·2	136·825	63·175
etc.	etc.	etc.	etc.	etc.	etc.

The results can be obtained graphically as follows. Prepare the axes of the graph, as in fig. 25, shewing voltage up to 200 volts and time up to, say $\cdot 1$ second. From the origin of the graph draw a straight line to the point $E = 200$, $t = CR = \cdot 01$ second. This shews the rate at which the counter-E.M.F. of the condenser would increase if the initial charging current were maintained. It is seen that the P.D. of the condenser plates after $\cdot 0025$ second would be 50 volts, as calculated above. The difference between the applied and counter-E.M.F. (150 volts) will now tend to cause the condenser P.D. to rise from 50 to 200 volts in $\cdot 01$ second, and therefore a line is drawn from the point 50 volts, $\cdot 0025$ second, to the point 200 volts, $\cdot 0125$ second, i.e. $\cdot 01$ second further along the time axis, but on the 200 volt line. At $\cdot 005$ second, this line gives the condenser P.D. as 87·5 volts, as in the table. The complete curve shewing the increase of condenser P.D. can be constructed in this manner, drawing in each successive "voltage increase" line to a point on the 200 volts ordinate, a time interval of $\cdot 01$ second ahead of the previous line. The graphical process should be performed by the reader for other values of the circuit constants, as the construction has many other applications, for instance in Chapter II we find the growth of current through an inductive circuit in the same way. It must however be pointed out that the curve obtained in the figure is not quite accurate, because the time interval, $\cdot 0025$ second, during which the charging current is assumed to remain constant at each successive value, is much too long. The shorter this interval is assumed to be, the greater will be the accuracy of the graphical construction. It can be proved that the current will reach $\cdot 632$ of its maximum value in a time equal to the time constant of the circuit. By the graphical construction given above, the voltage rises to 136·8, which is $\cdot 685$ of the maximum value, in a time equal to the time constant. If the intervals are halved, the graphical construction gives a curve which rises to $\cdot 64$ of the maximum value in the time $T = CR$, i.e. $\cdot 01$ second, and is then a good approximation to the true curve; in fig. 25 several points on the latter are indicated by small circles. The reason for taking the longer time interval in the above example is simply to avoid crowding the lines on the graph.



METHOD OF PLOTTING CONDENSER VOLTAGE WHEN CHARGED AND DISCHARGED

Discharge of condenser through resistance

60. In fig. 25 the condenser P.D. is practically equal to the applied voltage after an interval of $\cdot 05$ second from the instant of closing the switch. If the switch S_1 is opened, the condenser would remain charged to this voltage if its insulation were perfect. As there is no such thing as a perfect insulant, the charge will gradually leak away, and the lower the insulation resistance between the plates, the more rapid will be the discharge. Let us suppose that we expedite the discharge by connecting the 1,000 ohm resistance across the terminals of the condenser by closing the switch S_2 . A current commences to flow as soon as the circuit is completed, its value, momentarily, being $\frac{V}{R}$, V being the P.D. between the plates and equal to E , the voltage of the charging battery. The quantity of electricity which will pass in the short interval t seconds will be $\frac{V}{R} t$, and the fall in P.D. will be the quantity divided by the capacitance, or $\frac{V}{CR} t$. If this rate of discharge could be maintained, the condenser voltage would fall to zero in a time T , where $\frac{V}{CR} T$ is the total fall of P.D., i.e. V volts. The time T is therefore CR seconds, which it will be remembered is the time which would be taken to charge the condenser to the voltage of the supply if the initial rate of charge could be maintained. We now see that the initial rate of discharge is equal to the initial rate of charge provided the resistance of the circuit is the same in both charging and discharging. The initial rate of discharge can be shewn on the graph by a straight line drawn from the point $t = \cdot 05$, $E = 200$ volts to the point $t = \cdot 06$ second, $E = 0$. The discharging process is very similar to the process of charge, in that after a short interval of time t , the voltage available to cause current to flow is not V , but $V - \frac{V}{CR} t$. When $t = \cdot 0525$ second, the voltage available to cause further current to flow is only 150 volts, which should be compared with the voltage available for charging in the previous example. This 150 volts will now tend to discharge completely the condenser in a further $\cdot 01$ second, and so a new "voltage fall" line may be drawn from the point $E = 150$ volts, $t = \cdot 0525$, to the point $E = 0$, $t = \cdot 0625$. Continued repetition of this process gives the discharge curve of the condenser. Instead of falling to zero in a time CR or $\cdot 01$ second, the P.D. falls to 36.8 per cent. of its original value. The rate of charge and discharge of a condenser when associated with a resistance becomes of practical importance in certain radio instruments, for example, the "grid condenser and leak" commonly found in valve transmitting apparatus.

Dielectric strength

61. (i) If the P.D. applied to a condenser exceeds a certain limit the strain in the dielectric becomes so great that the atoms composing it are forced to allow a conduction current to flow, and the dielectric is then said to be punctured. In a solid dielectric the puncture actually takes the form of a hole which is burnt through the dielectric, and a plate of insulating material which has suffered in this way is of no further service. If the construction of the condenser is such that the defective plate cannot be removed, the whole condenser is rendered unserviceable by the failure of the one plate of dielectric, hence it is important that the rated safe voltage of any condenser shall not be exceeded. In liquid and gaseous dielectrics, the insulating substance closes round the puncture, and such substances are said to be self-sealing, but it must be appreciated that in a liquid dielectric such as oil the phenomenon of puncture will be accompanied by carbonisation of the oil and its insulating properties will be impaired. Puncture of liquid or gaseous dielectrics is facilitated by the presence of foreign bodies of poor insulating qualities, e.g. dust, metallic particles or fluff, and great care should be exercised in the exclusion of such substances when assembling or reassembling any condenser.

(ii) The dielectric strength of any material is defined as the voltage required to puncture a plate of the material one mil. ($\cdot 001$ in.) in thickness, the types of electrode used to apply the voltage to the dielectric itself being specified. For reasons into which it is unnecessary to enter, the dielectric strength between sharp points is less than between flat plates, or than between

CHAPTER I.—PARAS. 62-65

spherical balls, and condenser dielectrics are usually tested under the conditions in which they will be used, i.e. with flat metallic electrodes between which the dielectric is clamped. A thin sheet of dielectric is found experimentally to be proportionately stronger than a thicker one of the same material. The reason for this is obscure but appears to be associated with the fact that as the thickness of the material increases, the electric field strength in the dielectric departs to a greater degree from uniformity, and in certain places may reach a higher value than is calculated on the assumption of uniform field strength. The dielectric strength of some common insulating materials is given in Table II, Appendix A. It must be borne in mind that dielectrics puncture at lower voltages if subjected to alternating E.M.F. than if subjected to steady E.M.F. and the higher the frequency of the applied voltage the lower is the voltage at which a given dielectric will puncture. Hence condensers are generally marked with their test voltage, its nature whether alternating or steady being also stated. Thus a condenser marked "4,000 volts, D.C. test", means that it has passed a test in which a steady voltage of 4,000 volts was applied for a considerable period. The latter stipulation is important, for if the dielectric allows a small conductive current to flow its temperature will increase and the dielectric may be weaker at the higher temperature than at the lower. This point must also receive due attention in applying a condenser to any particular purpose. If the charging and discharging currents are very high, the metal plates forming the condenser may become heated, and the consequent heating of the dielectric may cause it to fail at voltages below the test voltage.

INSULATING MATERIALS

62. The insulating materials used in general electrical work may be either solid or liquid, the former comprising hygroscopic materials such as fibre, paper, asbestos, etc., and non-hygroscopic materials such as rubber, mica, glass, etc. The liquid insulators are practically confined to oils which are hygroscopic, and varnishes which are non-hygroscopic after baking at a high temperature. The following brief notes deal with some of the principal characteristics of various insulants.

Non-hygroscopic substances

63. *Rubber*.—This substance is a complex vegetable hydrocarbon occurring in the natural form as a sticky mass called latex, which is obtained from certain trees in the equatorial zone. The natural product is converted into rubber by first evaporating the moisture from the latex in a wood fire while stirring with a wooden paddle. The latex is then dried and hardened, and subjected to a purifying process, which consists of soaking in hot water for a long period, slicing the softened rubber, and then washing and drying, the resulting product emerging as sheets of crepe rubber, which is dried in a dark room at a temperature of about 130° F. The density of rubber is from 0.92 to 0.96, and its specific resistance at 24° C. is 11×10^{15} ohms per cm. cube. By mixing with from 2 to 3 per cent. of sulphur under the application of heat, a process which is termed vulcanisation, a material called vulcanised rubber is produced, and this material is not affected by temperature changes to the same extent as pure rubber. A slight increase in sulphur content results in the production of hard rubber or ebonite.

64. *Gutta percha*.—This material is also the coagulated sap of certain trees, and resembles rubber in some respects. It differs however in the following characteristics :—

- (i) It softens at a low temperature of about 65° C.
- (ii) It is unaffected by immersion in water provided it is screened from light.

The latter property suggests its principal use, which is the insulation of submarine cables. The density of gutta percha is about 0.98 and its specific resistance 2×10^9 ohm per cm. cube.

Chatterton's compound, which is used extensively in cable repair work, is made from the following materials. Gutta percha 60 per cent., Stockholm tar 20 per cent, resin 20 per cent.

65. *Mica*.—Mica is one of the most important electrical insulators, having a high dielectric strength and being capable of withstanding extremely high temperatures. It has certain disadvantages, however, being mechanically weak and only obtainable in thin sheets. It is a

mineral and is mined in India, Canada and U.S.A., being found in the crevices of certain igneous rocks, and its high price is in part owing to this, for often one ton of rock must be removed in order to obtain one pound of mica. The varieties of mica known as amber, green and ruby are slightly different in composition and the ruby is most used for high class work. Mica is not easily corroded and it withstands most acids and alkalies, but oil penetrates between its laminae in course of time and causes disintegration. The laminated structure is the outstanding feature of this substance, no other mineral possessing it.

The principal uses of mica for electrical purposes are :—

(i) Insulation of commutator segments (*see* Chapter IV). Canadian amber mica is used for this purpose, because under friction against a carbon brush it wears at the same rate as copper.

(ii) Insulation of heating units in apparatus such as electric soldering irons.

(iii) Manufacture of condensers for radio purposes, and also for magnetos. Ruby mica is invariably used for these, only the highest quality being accepted for transmitting condensers or magneto condensers.

(iv) Insulation of central electrode of sparking plugs for petrol engines. (*See* A.P. 1464, Engineering Manual.)

Micanite is extensively used in tubular form, e.g. in the slots of the armatures of dynamo electric machinery. (Chapter IV.) It is made by pasting together with shellac-varnish layers of mica flakes, sometimes with the addition of thin paper or cloth.

66. *Porcelain* is an artificial product, composed of china clay, flint and other ingredients which are ground to a fine powder with water, forming a plastic substance which is moulded to shape and afterwards baked. There is considerable shrinkage during the latter process, and it is not adapted for work which requires fine limits. It has no superior for such purposes as transformer terminal insulators, and for the insulation of overhead wires.

Marble and *slate* are good insulators possessing considerable mechanical strength, and like porcelain they are incombustible. Their chief use is in large switch-board panels.

Shellac is a natural product, the deposit of a certain kind of insect. It is a reddish brown substance which is sold in flakes, but is invariably used as a varnish which consists of flake shellac dissolved in alcohol.

Hygroscopic substances

67. *Paper* is extensively employed for insulating cables, being wound round the conductor in spiral form, and afterwards varnished and lead sheathed, the latter being almost compulsory owing to the hygroscopic nature of paper. In the manufacture of small parts such as magneto armatures and meter coils, a manilla paper coated with shellac is often used.

68. *Press-pahn* and *vulcanised fibre* are fibrous materials built up from laminae which are pressed together. Press-pahn is usually used in sheet form, while vulcanised fibre can be obtained in the form of sheet, rod or tube. Small bushes are often made of fibre in preference to ebonite on account of its greater mechanical strength, the bushes being frequently soaked in molten paraffin wax, which is an extremely good insulant, but is too weak mechanically for use except as a filling for such purposes, and for the construction of waxed-paper condensers.

69. *Asbestos*. This is a mineral which is found principally in Canada, South Africa and Italy. It is very hygroscopic, but can withstand extremely high temperatures. It is not used alone, except for the insulation of resistance elements, but forms an ingredient of several moulded compositions such as ebonestos and isolite.

Moulded compositions

70. The use of moulded parts has greatly increased of late years, owing to the reduction of cost compared with machined parts. The materials used generally consist of two principal ingredients, called the binder and the filler respectively. Moulding materials are classified according to the nature of the binder, the principal being rubber, natural resins and synthetic resins.

CHAPTER I.—PARAS. 71-74

71. *Ebonite*.—This is the best known rubber composition. It consists of crepe rubber with the addition of about 40 per cent. by weight of sulphur, the mixture being amalgamated under hot rollers. This product, which is called "dough", is pressed to the required shape in a steam-heated hydraulic press during which process partial vulcanisation occurs, the process being completed by "curing" the moulding in a chamber filled with steam. Ebonite cannot be moulded within fine tolerances, some machining being invariably necessary. Ebonite has the following disadvantages, viz. extreme brittleness and poor mechanical strength, it undergoes chemical decomposition under the influence of sunlight, and it softens at about 60° C. On the other hand, its dielectric strength is very high.

72. *Stabalite*.—This material contains rubber and sulphur with the addition of certain mineral ingredients, and is chiefly used in the manufacture of certain radio parts, and in magnetos, particularly the distributor of the latter. It has all the advantages of ebonite, but none of its drawbacks. It will safely stand temperatures up to 100° C.

The materials using natural resin, such as shellac, resin, asphalt or bitumen, as the binder, are known by many proprietary names; the fillers used are often wood pulp, magnesia, lime, sand and asbestos. None of these can be compared with ebonite or stabilite for dielectric strength and few will withstand greater temperatures than ebonite.

The synthetic resinous compounds consist of a binder which is formed by the action of phenol and formaldehyde, the fillers being wood pulp and asbestos. The best known of these products is bakelite which has good mechanical strength, high dielectric strength, and heat resisting properties, being efficient at temperatures up to 200° C. A further advantage is its excellent moulding properties, which have made it possible to mould items within a tolerance of a few thousandths of an inch. Other phenolic compounds are used for the panels of aircraft radio equipment.

Types of condensers

73. It was formerly usual to distinguish between "transmitting" and "receiving" condensers, but the necessity for this demarcation has disappeared to a large extent, because the peak voltages used in modern transmitters (other than those of very high power) are much lower than in the early days of radio communication, while on the other hand the efficiency of receiving condensers which formerly was not regarded as of very great importance, now receives a considerable amount of attention. A more useful classification is (i) low voltage, large capacitance, fixed value condensers, (ii) low voltage, small capacitance, fixed value condensers, (iii) high voltage, small capacitance, variable condensers. The main features and application of these classes will now be given.

74. Condensers of class (i) are commonly further subdivided into (a) non-inductive and (b) inductive types. The inductive type offers no electrical advantages and in some circumstances has positive drawbacks, but it is inexpensive and can be produced by automatic machinery. The dielectric is invariably of waxed paper, and carries a metallic coating which may be tin, lead or aluminium foil, but latterly is often applied to the dielectric by spraying, an alloy of low melting point being used. Two paper strips with their metal electrodes together with unmetallised paper strips for separating purposes are rolled upon a mandril, values of capacitance up to 10 μ F being obtained. This method of construction gives long thin electrodes, and the current density in the dielectric near the points to which terminal connection is made is greater than at points remote from the terminal connections. A condenser having this "rolled-up" form possesses an inherent inductance which is of the order of a few microhenries and in certain circuits this causes serious complications. The non-inductive type is preferable in all respects, but is more expensive to make, as it is built up of flat plates, interleaved with a suitable dielectric, which may be waxed paper but is more often thin mica. Even so, there must be some slight residual inductance, but this is much less than the inductance of a condenser of rolled up construction. Non-inductive condensers are made in all values of capacitance up to about 1 μ F, larger values than this being seldom required. It is important to observe that even a few inches of connecting wire may seriously prejudice the behaviour of such a condenser, but further consideration of this point must be deferred until later.

75. Condensers of class (ii) are of similar construction to the non-inductive type just discussed, but either mica or air is invariably used as the dielectric. When mica is adopted the condenser is often assembled in two, three or four sections which are connected in series, in order that the condenser will withstand the desired voltage. Copper foil electrodes are usually employed, and a range of capacitance between $\cdot 00005$ and $\cdot 005 \mu F$ can be obtained, the overall dimensions being quite small, e.g. a $\cdot 005 \mu F$ condenser for 500 volts D.C. test may occupy about 0.5 cubic inch. When extremely low dielectric losses are essential, and the desired capacitance does not exceed about $\cdot 0001 \mu F$, air dielectric may be employed; the construction then resembles an air dielectric variable condenser, except that no provision is made for alteration of capacitance. Condensers of class (iii) are commonly employed in the radio frequency circuits of both transmitters and receivers. The condenser consists of a number of fixed plates of brass or aluminium, which are in electrical connection and form one electrode of the condenser. A second set of plates, generally called the moving vanes, are capable of rotation about a central axis. These vanes are often of semi-circular shape, and rotation of the spindle upon which they are mounted causes them to mesh into the spaces between the fixed plates so forming a condenser, the capacitance of which can be altered by variation of the amount of overlap between the fixed plates and the moving vanes. Air dielectric is almost universal for these condensers although occasionally thin ebonite, synthetic resin or mica plates are inserted between fixed and moving vanes, and are generally free to rotate with the latter. The dielectric then consists partly of air and partly of solid material and the capacitance is somewhat increased. In the service Type 7 condenser, which is a variable condenser of $\cdot 0009 \mu F$ maximum capacitance normally, and is of very robust construction, the plates are enclosed in a glass vessel which can be filled with oil, if necessary. By this means the maximum capacitance can be increased to a value about three times its normal.

CHAPTER II.—MAGNETISM

PERMANENT MAGNETISM

Magnetic polarity

1. Everyone is familiar with the toy horse-shoe magnet, so called from its shape. It has the property of attracting iron filings, pins, needles and in fact, any small pieces of iron or steel. Such a magnet is called a permanent magnet because the attractive force possessed by it is inherent in the magnet itself and does not depend upon any external influence. Actually any permanent magnet slowly loses its magnetic properties, the degree to which they are retained depending upon the retentivity of the material of which the magnet is made.

A more convenient form of magnet for experimental purposes is the bar magnet, which can be made from an ordinary steel knitting needle by stroking it in one direction only with one pole of a permanent magnet. If such a bar magnet is suspended horizontally, as shewn in fig. 1a, it will be found to take up such a position that its axis lies approximately north and south. If one end is marked, it will be found that no matter how the magnet is displaced, the same end eventually comes to rest pointing roughly towards the north. The ends of the magnet are therefore called the north-seeking end (or "pole") and south-seeking end (or "pole") respectively.

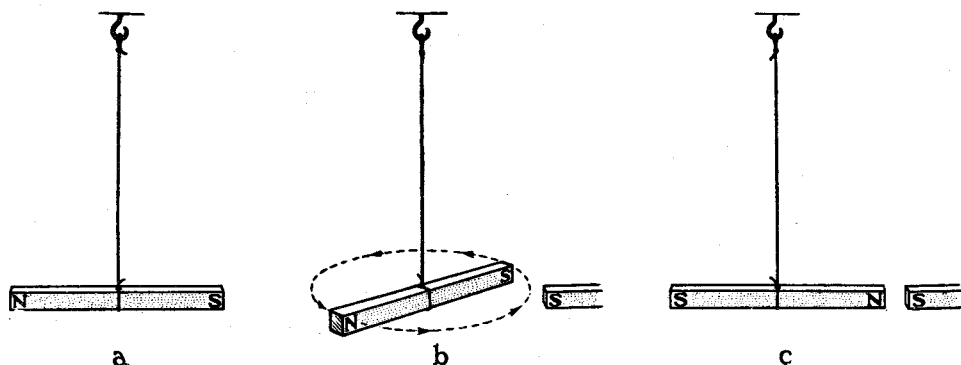


FIG. 1, CHAP. II.—Suspended bar magnets.

This pole-seeking property is due to the fact that the earth itself is a huge magnet, having south-seeking and north-seeking poles situated near, but not coincident with, the geographical poles. The north-seeking pole of a magnet is that which is attracted towards the geographical north region of the earth, but it is impossible to allot north-seeking or south-seeking properties to the earth itself for north or south has no significance except in relation to that body. If instead of the terms north-seeking and south-seeking, the poles are considered to be positive and negative respectively, the geographical north region of the earth is of negative polarity. It is desirable to avoid the use of the terms "north-pole" and "south-pole" when referring to magnets, substituting "N-pole" and "S-pole", which signify "north-seeking" and "south-seeking" poles. The attractive force of a magnet is most concentrated in the neighbourhood of its poles. This is easily shewn by dipping a magnet into iron filings, and noting in which regions most filings adhere.

The mutual action which takes place between magnets can easily be demonstrated, e.g. if the south-seeking pole of another magnet is brought near to the south-seeking pole of the suspended one, as in fig. 1b, it will be found that repulsion takes place, while if the north-seeking pole is presented to it attraction occurs, fig. 1c. The first law of magnetism is that "like poles repel, and unlike poles attract each other"

CHAPTER II.—PARA. 2

2. If a bar magnet is laid upon a table, and a small compass needle—which is merely a pivoted bar magnet—is moved about from point to point in its vicinity, the needle will be found to vary

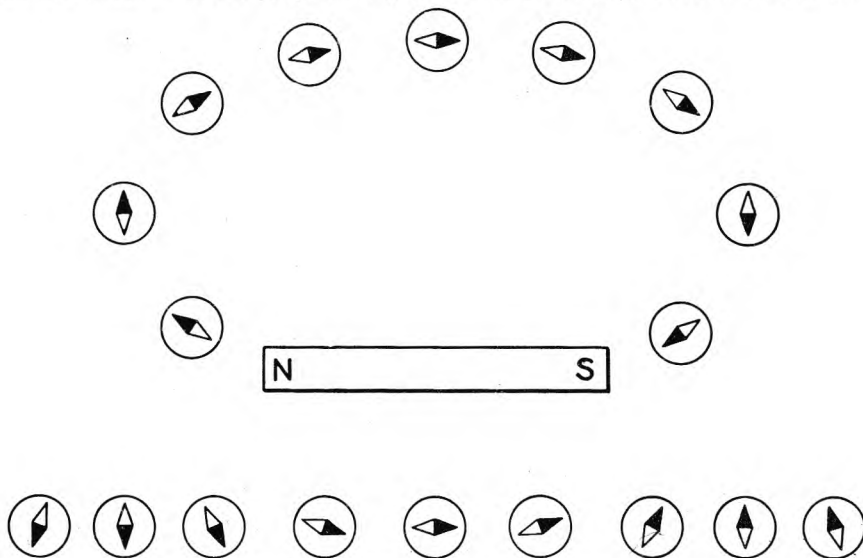


FIG. 2, CHAP. II.—Compass needles in magnetic field.

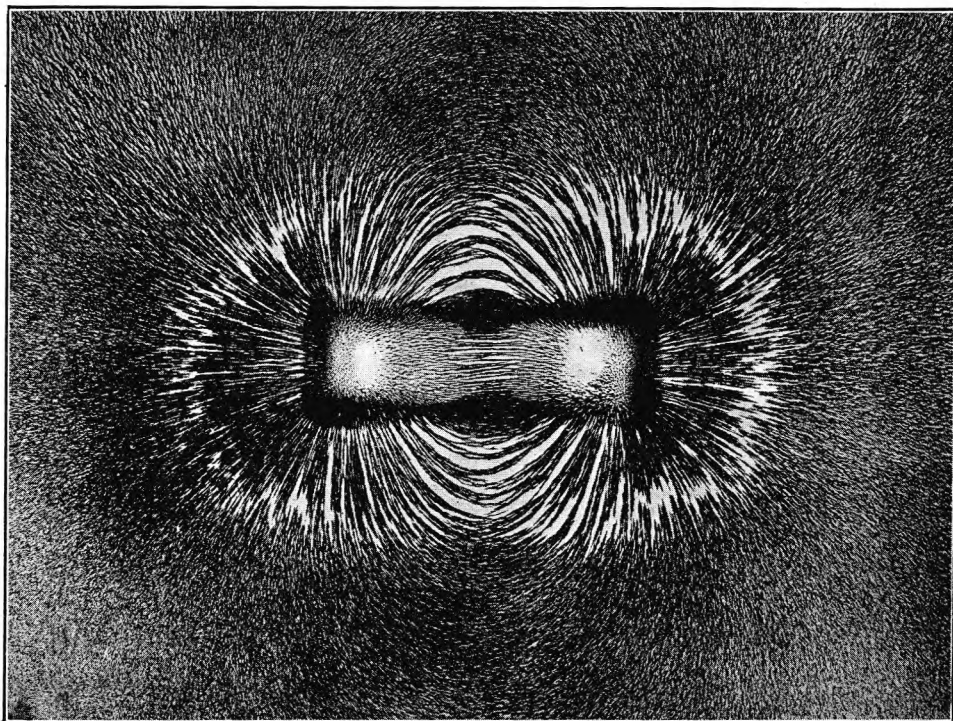


FIG. 3, CHAP. II.—Lines of force shown by iron filings around magnet.

its direction in each individual position. Fig. 2 shews the various directions taken up by the needle in such an experiment. The needle shews the direction of the magnetic force at the point at which it is situated. The magnetic field of a magnet is the region in which its action can be observed, and this field may be said to consist of lines of magnetic force, a line of force being defined as the imaginary line along which the force of a magnet acts. Faraday's conception of the properties of electric lines of force was explained in Chapter I, and it may be assumed that magnetic lines of force behave in an identical manner.

By a long-standing convention the positive direction or sense of the lines of force is taken as that direction in which a north-seeking pole would be urged if it were free to move. Thus it may be considered that in the external magnetic circuit the lines of force flow from north-seeking to south-seeking poles externally, and from south-seeking to north-seeking poles inside the magnet, each magnetic line being a complete closed loop. In diagrams an arrow head is usually placed upon some of the lines to indicate the sense of the field. From this idea of "flowing" we are led to speak of the magnetic flux of a magnet, as synonymous with the total magnetic field.

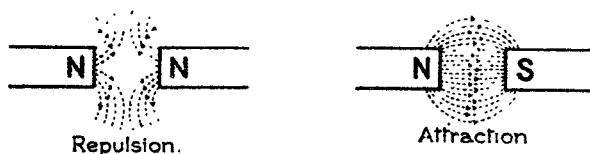


FIG. 4, CHAP. II.—Lines of force between adjacent magnetic poles.

3. A map or picture of the magnetic field in any one plane can be made by sifting fine iron filings from a small sieve on to a sheet of tracing paper under which a magnet has been placed. The filings then align themselves in the direction of the lines of force. It will be found advantageous to tap the sheet gently as the filings fall on it. Fig. 3 is reproduced from a photograph of the magnetic field of a magnet, which was produced in this way, while fig. 4 shews qualitatively the distribution of the field in the neighbourhood of (i) two unlike poles and (ii) two like poles, placed adjacent to each other.

Theory of molecular magnetism

4. The process by which a piece of iron or steel becomes magnetised has been explained as follows. The molecules of such substances are themselves minute magnets. In the non-magnetised state these molecular magnets align themselves into closed magnetic chains as shewn in fig. 5, in which the molecules are represented by small rectangles, the black portion being of north-seeking and the white portion of south-seeking polarity. Each little chain forms a

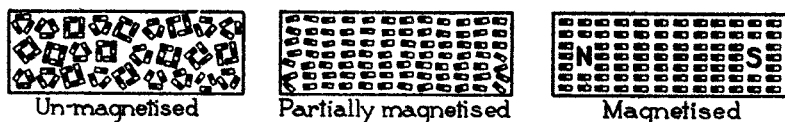


FIG. 5, CHAP. II.—Arrangement of molecules in magnetic substance.

complete magnetic circuit and no force is exerted externally. When the substance is magnetised, however, these molecular magnets are dragged into alignment along the axis of the magnet, the lines of force associated with them being now completed through the space surrounding the iron. The shape of the magnetisation curve of iron (*see* para. 22) strongly supports this theory. The difference in the behaviour of iron and steel in this respect compared with most other materials is not yet understood.

Induced magnetism

5. If a piece of soft iron be placed in the field of a permanent magnet, and the resulting field plotted by the iron filings method, the result is as shewn in fig. 6. It will be seen that the

CHAPTER II.—PARA. 6

soft iron now appears to have a magnetic field of its own and is said to be magnetised by induction. Consideration of the polarity of the permanent magnet and the soft iron shews that the adjacent poles are unlike and will therefore attract each other. It is in this way that all attraction between magnets and unmagnetised bodies commences, and may be summed up in the statement that "induction always precedes attraction".

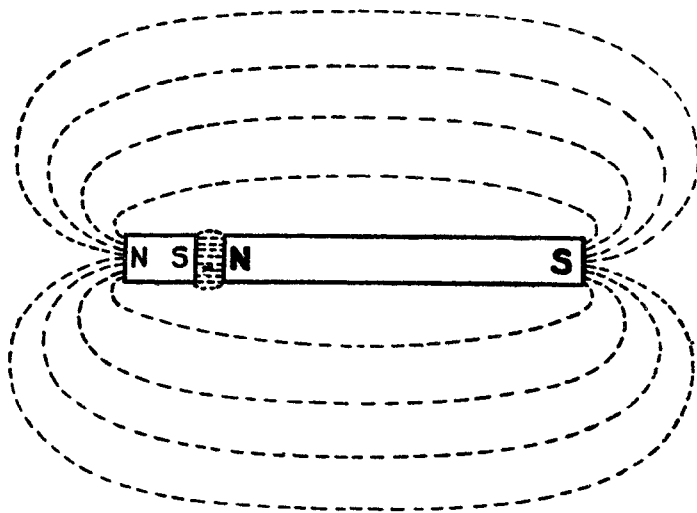


FIG. 6, CHAP. II.—Induced magnetism in soft iron.

ELECTROMAGNETISM

6. In the preceding paragraphs only the phenomena associated with or due to permanent magnets has been considered. The connection between magnetism and electricity was suspected for many years before it was finally established that an electron in motion always produced a magnetic field. Since an electric current consists of a motion of electrons, or possibly a motion of positive and negative ions, each of the latter possessing one or more surplus electrons, we may say that an electric current produces a magnetic field. This field forms circles concentric with the axis of the current-carrying conductor. It is important in certain circumstances (*see* Chapter V) to remember that this field exists both inside and outside the conductor. It can be demonstrated by the use of iron filings as in the case of a permanent magnet, and is shewn diagrammatically in fig. 7a. The direction of the field is shewn conventionally by the arrows, and it will be seen that the direction of the current and the direction of the magnetic field are related in the same way as the thrust and turn respectively of the ordinary corkscrew. This useful mnemonic is usually known as "Maxwell's corkscrew rule".

An immediate practical application of this rule is to indicate the direction of current in a wire. If a conductor carrying a current is laid over and parallel with the compass needle, the needle will be deflected, tending to place itself at right angles to the conductor. The actual angular deflection in any particular case will depend upon the strength of the magnetic field due to the current, and to the controlling force of the earth's magnetic field. If the direction of the current in the wire is from south to north, the north-seeking pole of the needle will be deflected to the left, i.e. toward the west, while if the direction of the current is reversed, the deflection will be toward the east. This is most easily remembered by Ampere's "swimming rule":—

"Consider a man swimming face downwards with the current in the conductor, over the compass needle. The north-seeking pole of the latter will be deflected towards his left."

7. If two wires carrying currents are placed parallel to each other the resulting magnetic field may be depicted as in figs. 7b and 7c. In fig. 7b the currents are in opposite directions, and in fig. 7c they are in the same direction. The conductors in the former case tend to attract and in the latter to repel each other. An experiment to show this can easily be performed, the apparatus being arranged as in fig. 8. Two light metal rods are suspended from a suitable support,

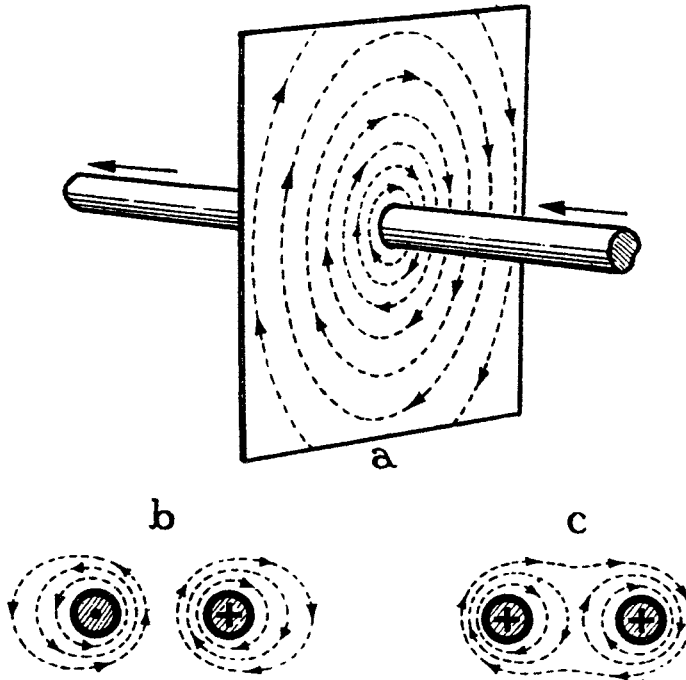


FIG. 7, CHAP. II.—Magnetic fields around conductors.

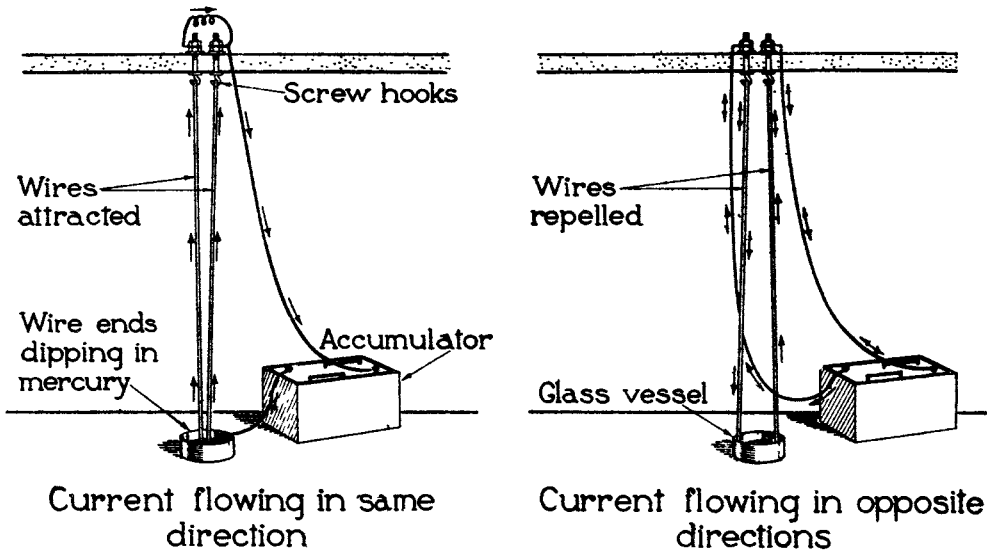


FIG. 8, CHAP. II.—Magnetic action of parallel currents.

CHAPTER II.—PARAS. 8-9

which serves also to insulate them from each other. The lower ends of the rods dip into a vessel containing mercury, and electric current is supplied by means of a battery. By making suitable connections, the currents may be made to flow either in the same or in opposite directions in each wire, and the resulting attraction or repulsion may be observed by the creation of ripples upon the surface of the mercury. The magnetic fields caused by the current in each wire tend to remain concentric with the conductors, and the latter move in order to allow this. In practice the foregoing effect has to be allowed for when designing parallel conductors carrying very large currents, e.g., the "bus bars" in a large power station.

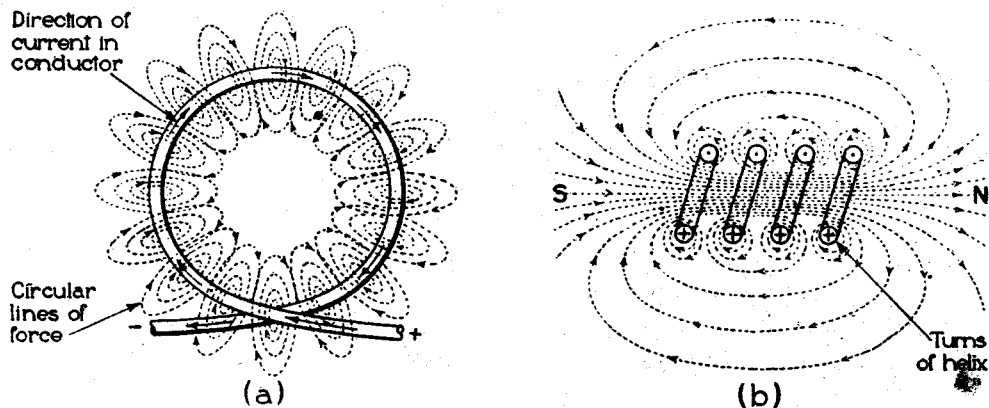


FIG. 9, CHAP. II.—Magnetic fields around coiled conductors.

8. If the current carrying conductor is bent into a single loop as in fig. 9a the relative configuration of its magnetic field will not be altered, but all the lines of force will now enter one face of the loop and emerge on the other. Several continuous loops wound in this way, forming a spiral or helix are said to constitute a solenoid. The configuration of the magnetic field of a solenoid is shown in fig. 9b, and it will be seen that the resultant field is similar to that of a bar magnet. All the phenomena associated with a bar magnet can, in fact, be reproduced equally well by the solenoid. The magnetic polarity of a solenoid depends upon the direction of current round the windings, and is generally found by the following rule. "Looking at one end of the coil, if the current flows in a clockwise direction, the end nearest the observer is a south-seeking pole. If current is in anti-clockwise direction, the nearest end is a north-seeking pole." The appropriate mnemonic for this rule is shown in fig. 10.

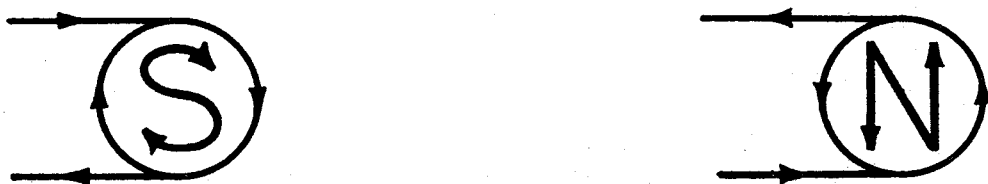


FIG. 10, CHAP. II.—Mnemonic for polarity of solenoid.

The galvanometer

9. The deflection of a magnetic needle by the current in a single conductor will be increased if the current is increased, or if the number of current-carrying conductors affecting the magnet is increased, provided that these conductors are so arranged that their effect is cumulative. This is easily achieved by winding a short coil of many turns of wire on a suitable "former," the magnetic needle being pivoted in the centre. Such an arrangement is called a galvanometer.

By applying Ampere's Swimming Rule, it will be observed that the current in the portions of conductor above the needle tend to deflect the latter in the same direction as those beneath the needle, so that by using a great many turns of fine wire, very small currents may be detected. The essential portions of this type of galvanometer are shown in fig. 11. Other, and more sensitive, types of current indicator will be discussed in the chapter devoted to Measuring Instruments.

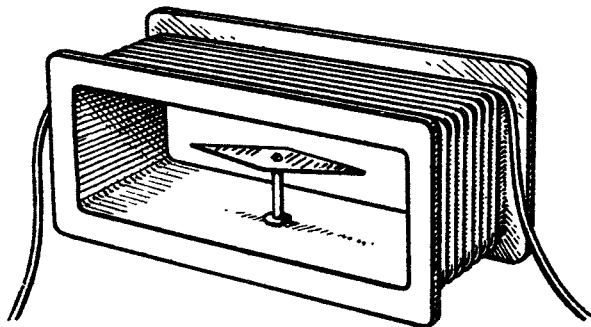


FIG. 11, CHAP. II.—Galvanometer.

THE ELECTRO-MAGNETIC SYSTEM OF UNITS (E.M.U.)

Magnetic field strength

10. If two like poles of equal strength, situated 1 cm. apart in vacuo, repel each other with a force of one dyne, they are said to be unit poles, or to possess unit pole strength.

The force exerted between such poles at any distance is inversely proportional to the square of the distance between them. This is the Second Law of Magnetism and was demonstrated by Coulomb. Suppose an isolated pole has n lines of force associated with it, and is surrounded by a spherical surface of r cm. radius, the pole being at its centre. Since the area of this surface is $4\pi r^2$ sq. cm. the number of lines of force passing through each square centimetre will be $\frac{n}{4\pi r^2}$. If another concentric spherical surface is situated at a distance of $2r$ centimetres from the pole, the number of lines passing through each square centimetre will be $\frac{n}{4\pi (2r)^2}$ or $\frac{n}{16\pi r^2}$. Thus at twice the distance from the pole, only one-fourth of the lines of force are effective in any given area.

The strength or intensity of a magnetic field at any point is measured by the force exerted on a unit pole, if placed at that point, and is denoted by the symbol H . Magnetic field strength is measured in dynes per unit magnetic pole, thus a pole of strength m units sets up at a distance d cms. from it in a vacuum a magnetic field strength of $\frac{m}{d^2}$ dynes per unit magnetic pole. The name "oersted" is sometimes used for a field strength of one dyne per unit pole.

Magnetic Flux

11. The total number of lines of force flowing from the north seeking or N-pole to the south seeking or S-pole has already been referred to as the magnetic flux. The lines of force bounding any area in the field perpendicular to its direction may be considered to enclose a tube of flux. From a unit magnetic pole it is assumed that one unit tube of flux passes through each square centimetre of the surface of a sphere one centimetre in radius, having the pole as its centre. The surface area of this sphere being 4π square centimetres it follows that 4π unit tubes of flux are assumed to emanate from or terminate upon a unit pole. The unit of magnetic flux is the unit tube or Maxwell, the symbol for magnetic flux being Φ .

CHAPTER II.—PARA. 12-13

The moment of a magnet is the product of the pole strength and the distance between the poles. If the pole strength is m unit poles and they are l centimetres apart, the moment is ml units.

The intensity of magnetisation of a magnet is its moment per unit volume or $\frac{ml}{v}$. If the magnet is of uniform cross section A the volume v will be lA , and the intensity of magnetism which is denoted by J will be $\frac{ml}{lA}$ or $\frac{m}{A}$. Hence the important result that the intensity of magnetisation is synonymous with the pole strength per unit area.

Flux density is defined as the number of unit tubes of flux passing through unit area in the field perpendicular to its direction. In a vacuum a magnetic field of strength H oersteds sets up a magnetic flux density of B_0 tubes per square centimetre, B_0 being numerically equal to H although its unit is different. In a uniform magnetic field of area A and strength H oersteds the total magnetic flux is B_0A tubes, which is numerically equal to HA . The name gauss is given to the unit of flux density, one gauss being equal to one unit tube per square centimetre.

12. If a piece of soft iron of uniform cross section is introduced into the uniform field just mentioned, the iron will be magnetised by induction, and it will develop magnetic poles upon (or near) each surface perpendicular to the field. Suppose these poles each to be of pole strength m units, then the intensity of magnetisation upon each of these surfaces will be $J = \frac{m}{A}$. The soft iron now has its own field, the flux through the iron due to its own magnetism being 4π times the pole strength at either end, and the total flux will therefore be B_0A due to the magnetising agency and $4\pi m$ due to the induction, or, putting for the total flux Φ ,

$$\Phi = 4\pi m + B_0A$$

The flux density in the iron, $B = \frac{\Phi}{A}$

$$\begin{aligned} B &= \frac{4\pi m}{A} + B_0 \\ &= 4\pi J + B_0 \\ &= B_0 \left(\frac{4\pi J}{B_0} + 1 \right) \end{aligned}$$

Since B_0 is equal numerically to H , the field strength causing the magnetisation,

$$B = H \left(\frac{4\pi J}{H} + 1 \right)$$

The factor $\left(\frac{4\pi J}{H} + 1 \right)$ is called the permeability of the material in which the field is situated, and is denoted by μ . The relation between B and H is therefore written $B = \mu H$.

Permeability

13. (i) The permeability of a medium may be defined as the numerical ratio of the flux density to the magnetic field strength, or alternatively as the ratio of the flux density in the medium to the flux density in vacuo for the same value of magnetic field strength. The absolute value of the permeability is not known, but it is assumed to be unity in vacuo, and the value for other materials is then only comparative. According to the nature of their permeability, materials are divided into ferro-magnetic, dia-magnetic and para-magnetic materia. Ferro-magnetic materials are those whose permeability is large compared to unity. A characteristic of these substances is that μ itself is not constant but varies according to the flux density, the nature of this variation being shown by the slope of the magnetisation or B/H curve of the material,

as shown in fig. 12. Iron, steel, nickel and cobalt and many, but by no means all, of their compounds are ferro-magnetic, and included in this group are the manganese bronzes discovered by Heusler in 1898, which have magnetic properties somewhat resembling those of cast iron.

(ii) Para-magnetic substances are those which have a permeability which is constant for all values of flux density and is only slightly greater than unity. They behave magnetically like iron, but to a much smaller degree, while dia-magnetic substances behave in an entirely opposite manner, the permeability being constant for all values of flux density, but slightly smaller than unity. The most dia-magnetic substance known is bismuth, and a ball of bismuth is not attracted by a magnet but repelled, while a bar of bismuth placed in a magnetic field sets itself in a transverse direction instead of in line with the direction of the field as a ferro-magnetic bar would do. The separation of materials into para-magnetic and dia-magnetic classes is not yet complete, since

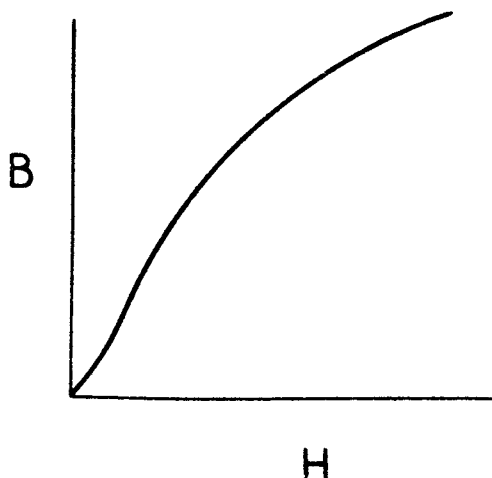


FIG. 12, CHAP. II.— B/H curve.

the measurements necessary are so delicate and usually have to be made in air (which is itself para-magnetic) in the presence of the earth's magnetic field. It is now believed that no substance is absolutely non-magnetic, and a recent tabulation of magnetic properties of various substances gives solid oxygen, manganese, iron oxide, platinum, chromium, tantalum, and aluminium as definitely para-magnetic; sodium, potassium, and wood as doubtful; copper, sulphur, glass, zinc, quartz, lead, silver, gold, mercury, and bismuth as dia-magnetic, the most para-magnetic materials being mentioned first, and the most dia-magnetic last. It must be again emphasised that in para-magnetics and dia-magnetics, the permeability only differs from unity by a few parts in a thousand, e.g. for solid oxygen, $\mu \doteq 1.0053$, while for bismuth $\mu \doteq .99983$. In ordinary engineering practice it is therefore usual to consider that all materials other than the ferro-magnetics have a permeability of unity.

Difference of magnetic potential

14. Consider any two points in space, A and B, and let a unit north-seeking pole be transferred from A to B in the presence of magnetic forces in the region between the two points. In doing this, a definite amount of work is performed, which is independent of the shape of the path along which the unit pole is moved. The amount of work (in ergs per unit pole) is the difference of magnetic potential between the points A and B. If work is done on the charge in moving from A to B, B is at a higher potential than A, while if work is done by the magnetic forces themselves, A is at a higher potential than B, that is a north-seeking pole tends to move from a point of high to a point of low potential.

CHAPTER II.—PARAS. 15-17

The electromagnetic unit of current

15. Since a magnetic field can be produced by an electric current, it is desirable that unit current should be capable of definition in terms of its magnetic effect. The electromagnetic unit of current is the current which, when flowing in a wire bent into an arc of a circle of 1 centimetre length and 1 centimetre radius, produces unit magnetic field strength at the centre of the arc. It may also be defined as the current which, flowing in a single circular turn of 1 centimetre radius, produces at the centre of the circle a magnetic field strength of 2π units. The E.M.U. of current is equal to 10 amperes, and since the unit of quantity is unit current $\times$ unit time, it follows that the E.M.U. of quantity is 10 coulombs. Having defined the E.M.U. of quantity the E.M. units of electric P.D. and resistance follow :—

(i) Unit P.D. is the P.D. between two points when the work done in carrying unit quantity from one point to the other is 1 erg.

$$\begin{aligned} 1 \text{ unit of P.D. (E.M.U.)} &= \frac{1 \text{ erg}}{1 \text{ unit of Quantity (E.M.U.)}} \\ &= \frac{1}{10^7} \frac{\text{joule}}{\text{10 coulomb}} \\ &= \frac{1 \text{ joule}}{10^8 \text{ coulomb}} = \frac{1}{10^8} \text{ volt.} \end{aligned}$$

$\therefore$ 1 volt = 10^8 electromagnetic units of P.D. or E.M.F.

(ii) Unit resistance (E.M.U.)

$$\begin{aligned} 1 \text{ ohm} &= \frac{1 \text{ volt}}{1 \text{ amp.}} = \frac{10^8 \text{ E.M.U. P.D.}}{10 \text{ E.M.U. current}} \\ &= 10^9 \text{ electromagnetic units of resistance.} \end{aligned}$$

Magnetic field strength at the centre of a current-carrying loop

16. Consider a conductor carrying a current of I E.M. Units, bent into a circular loop of radius r centimetres. Then the magnetic field strength set up at the centre, by a very short element of the conductor of length δl , is δH , and $\delta H = \frac{I \delta l}{r^2}$. The total field strength H is the sum of all such elements of field strength due to the whole length of the conductor, which is $2\pi r$ centimetres, and therefore $H = \frac{I}{r^2} \times 2\pi r = \frac{2\pi I}{r}$ dynes per unit magnetic pole. If the current is in amperes, then $H = \frac{2\pi I}{10r}$ dynes per unit magnetic pole, or oersteds.

Force exerted upon a conductor in a magnetic field

17. (i) If a unit pole is placed at the centre of a circular loop, which is carrying a current of I E.M. Units there will be a mutual force between the unit pole and the equivalent pole producing the magnetic field of strength H . Since this equivalent pole is the conductor itself, the force will be radial, and therefore perpendicular to the length of the conductor. This result is perfectly general, the force between a magnetic field and a conductor always tending to urge the conductor in a direction perpendicular to its length and also perpendicular to the lines of force. A straight conductor of length l centimetres, carrying a current of I E.M. Units, when placed in a uniform field of strength H dynes per unit magnetic pole, will experience a force of $HI l$ dynes in a direction mutually perpendicular to both conductor and field.

(ii) The direction in which the conductor tends to move may be deduced from a consideration of the elastic properties of the imaginary lines of force. Fig. 13 shews a conductor situated in a magnetic field, and carrying a current which is assumed to flow out of the paper towards the reader. At (a) the magnetic field due to the current is shewn superimposed upon the original field, and it will be observed that on the left hand side of the conductor the lines of force are in the same direction, so that in this region the effect of the current is to strengthen the field, while on the right hand side of the conductor the field of the latter is in opposition to the original field and tends to reduce the field strength on this side. In reality the two fields of the magnet and conductor respectively combine to form a resultant field as indicated at (b). Owing to the repulsive action of parallel lines of force acting in the same direction, the distribution of the flux tends to become uniform, and if the conductor is free to move it will be displaced to the right, which is the region in which the field is weakest. The action may be attributed to the tendency of the lines composing the distorted magnetic field on the left of the conductor to

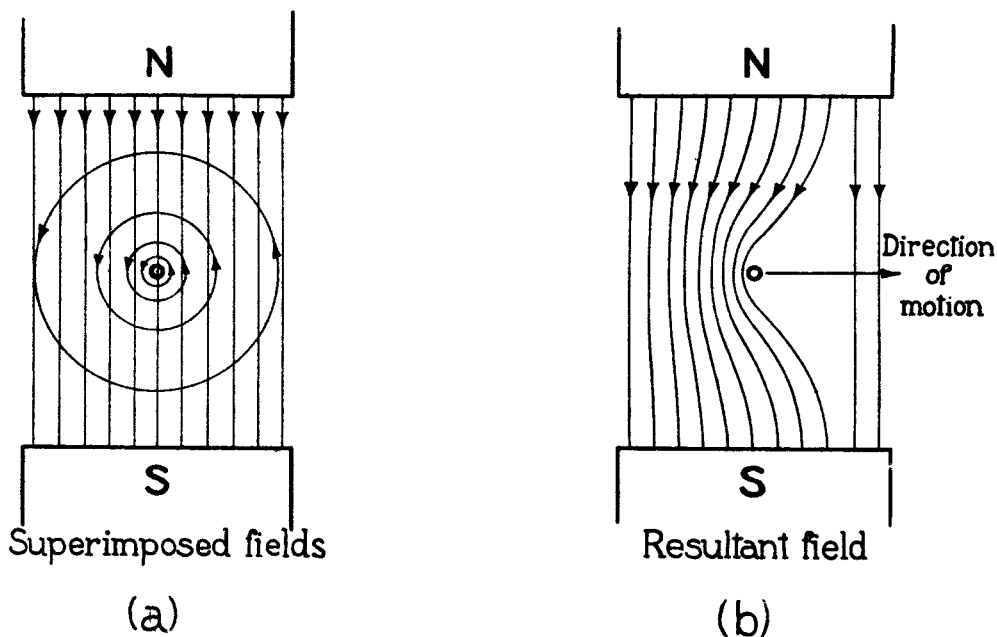


FIG. 13, CHAP. II.—Movement of current-carrying conductor in magnetic field.

shorten as much as possible and also to repel each other, so that the field tends to resume its original uniform distribution. The direction in which the conductor will move may be remembered by Fleming's Left Hand Rule, i.e. extend the thumb, forefinger and middle finger of the left hand in mutually perpendicular directions. Place the Forefinger in the direction of the Field, the middle finger in the direction of the current (**I**) and the thumb then indicates the direction of Motion of the conductor.

Work done by change of flux

18. If a conductor of length l centimetres carrying a current of I E.M.U. is placed in a field of strength H dynes per unit magnetic pole, the force acting on it, from the above discussion, is HI dynes. Now let the conductor be moved in opposition to this force through a distance of d centimetres, the work done being $HIld$ dyne-cms. or ergs, $= W$. Now H is numerically equal to the flux density, if the permeability of the medium is unity, ld is the area A swept by the conductor and HA is the change of flux linking with the electric circuit. The work performed

CHAPTER II.—PARAS. 19-20

is therefore equal to the product of the current and the change of flux, and this result is found to be true for any shape of electric circuit. Algebraically, $W = I (\Phi_2 - \Phi_1)$, Φ_2 being the flux enclosed before the conductor is moved and Φ_1 the flux enclosed after the operation has taken place.

The average rate at which this work is done, or the power expended, is

$$\frac{W}{t} = I \left(\frac{\Phi_2 - \Phi_1}{t} \right) \text{ ergs per second.}$$

Power being the product of E.M.F. and current,

it is apparent that the expenditure of mechanical energy has resulted in the production of an E.M.F.

Since
$$\frac{W}{t} = EI, EI = I \frac{(\Phi_2 - \Phi_1)}{t}$$

$$E = \frac{\Phi_2 - \Phi_1}{t} \text{ (E.M. units of E.M.F.)}$$

$$\text{or } E = \frac{\Phi_2 - \Phi_1}{10^9 t} \text{ volts.}$$

It should be observed that the creation of this E.M.F. is due to the conversion of mechanical energy into electrical energy. There has been no expenditure of magnetic energy although the magnetic flux played an important part in the transformation.

Magneto-motive force

19. The work done in carrying a unit pole round any closed path in a magnetic field is called the magneto-motive force in that path. First of all, let us suppose that the magnetic field is produced by a single circular turn of current-carrying conductor, and that the path of the pole does not encircle the conductor. Then if work has to be done on the pole in moving it towards the conductor, an equal amount of work is done on the pole by the magnetic forces in moving it away from the conductor, and in taking the pole round the complete path the total external energy supplied is zero. That is, the M.M.F. round a closed path not encircling a conductor is zero. If however a unit pole is carried round a path encircling a conductor, which carries a current of I E.M.U., every line from the pole will link with the circuit, the total number of tubes of flux linking with the conductor, will be 4π tubes (by definition of a unit pole). The M.M.F. round this path will therefore be $4\pi I$ ergs, and is independent of the shape of the path. Again, if instead of a single loop we have a solenoid of N turns, and the unit pole is taken round a path linking with all the turns, the work done will be $4\pi IN$ ergs. The number of flux linkages is equal to the actual flux linking with the conductor multiplied by the number of turns linking with the flux.

Field strength inside a toroidal coil

20. Imagine a long thin solenoid uniformly wound with an insulated conductor, upon a cylindrical rod and carrying N turns in all; after winding the coil may be removed from the rod and the ends of the coil brought round to meet each other so that the winding forms a ring of radius r cms. Such a coil is called a toroid, or toroidal coil, and is shewn in fig. 14.

Suppose a current of I E.M. Units is flowing through the conductor. If a unit magnetic pole is taken round the mean circumference of the ring threading each turn of wire in succession the work done will be $2\pi r H$ ergs, H being the field strength inside the coil, which is at present unknown. But $2\pi r H$ ergs is also the M.M.F. round the path, and the latter quantity has been shewn to be $4\pi NI$ ergs. Equating the two expressions

$$2\pi r H = 4\pi NI$$

$$H = \frac{4\pi NI}{2\pi r} = \frac{4\pi NI}{l}$$

where l is the length of path of the unit magnetic pole, or $2\pi r$ cms.

If we imagine further that the radius r increases without limit, any portion of the length of the winding may be considered as approximately a straight solenoid, and so for a solenoid whose length is large compared to its diameter the field strength at its centre is given by the expression

$$H = \frac{4\pi NI}{l} \text{ if } I \text{ is in E.M.U.}$$

$$\text{and } H = \frac{4\pi NI}{10l} \text{ if } I \text{ is in amperes.}$$

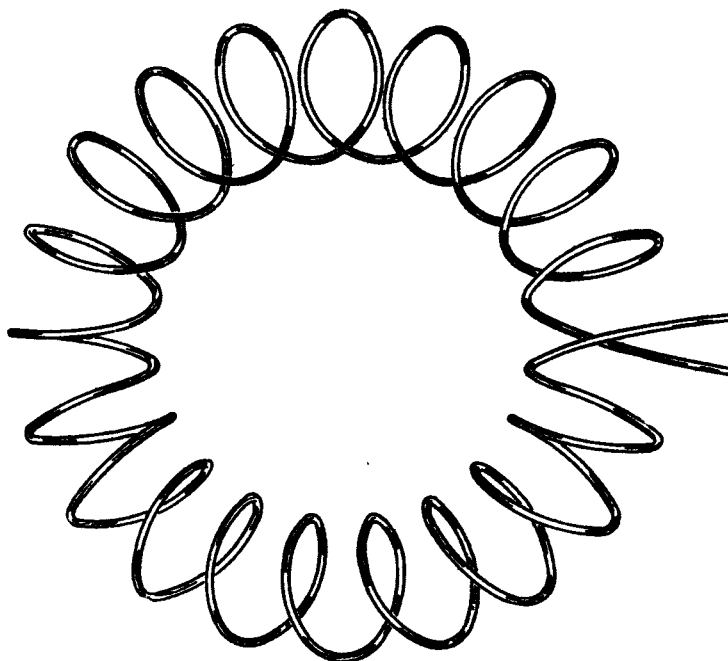


FIG. 14, CHAP. II.—Toroidal coil.

The quantity denoted by H is sometimes referred to as the magnetising force of the solenoid.

Suppose the space inside the winding to be filled with a substance of permeability μ . Then the flux density inside the material is B , or μH tubes per square centimetre, and the total flux is BA tubes.

$$\Phi = \frac{4\pi IN}{10l} \times \mu A \text{ maxwells.}$$

This equation is sometimes called Ohm's law for the magnetic circuit. It may be written

$$\Phi = \frac{\text{M.M.F.}}{\text{Reluctance}} = \frac{\text{M.M.F.}}{S}$$

$$\text{where M.M.F.} = \frac{4\pi IN}{10}$$

$$S = \frac{l}{A\mu} = \text{the reluctance of the magnetic circuit.}$$

The magneto-motive force (M.M.F.) is $\frac{4\pi}{10}$ (or 1.257) times the ampere-turns, and the reluctance of the magnetic circuit is the opposition offered by the substance to the establishment of a flux

CHAPTER II.—PARA. 21

It is thus analogous with the resistance of an electric circuit, i.e. its opposition to the establishment of a current. The appropriateness of the analogy is evident from an examination of the formula for the resistance of a conductor :—

$$R = \frac{l \rho}{A}$$

where l is the length of the conductor, A the area of its cross-section and ρ a constant for the material. Similarly, the reluctance of a magnetic path is given by

$$S = \frac{l}{A \mu}$$

Thus the permeability of the magnetic circuit in some respects resembles the specific conductivity of the electric circuit, but whereas the specific conductivity $\left(\frac{1}{\rho}\right)$ of a conductor is constant at uniform temperature, the permeability of a ferro-magnetic material varies with the flux density as explained later.

Example

Calculate the ampere-turns necessary to produce a flux of 10,000 lines in a closed iron ring of cross sectional area 4 sq. cm., μ being assumed constant and equal to 1000, and the mean length of a magnetic line in the iron being 20 cms.

Since
$$\Phi = \frac{1.257 \text{ } I N}{S}$$

$$I N = \frac{S \Phi}{1.257} = .8 S \Phi$$

$$S = \frac{l}{A \mu} = \frac{20}{4 \times 1000}$$

$$\begin{aligned} \therefore I N &= .8 \times \frac{20}{4000} \times 10,000 \\ &= \frac{8 \times 20}{4} = 40 \text{ ampere turns.} \end{aligned}$$

B. H. CURVES

The Thomson permeameter

21. In the last example, it was assumed that the permeability of the iron core was constant, and equal to 1000. In the case of ferro-magnetic materials, the permeability depends upon the quality of the iron, the flux density, and the temperature, as well as upon the previous magnetic history of the sample concerned.

The relation between B and H may be determined experimentally. If this relationship is plotted with H as abscissa and B as ordinate, the resulting graph is called the B/H curve for the particular sample. The Thomson permeameter is an early form of apparatus employed for such a determination, although more rapid methods are now generally used. The theory of the permeameter exhibits the quantitative relationship between B and H directly, and is chosen as an illustration for this reason.

The permeameter is shewn diagrammatically in fig. 15. It consists of a massive soft iron magnetic circuit or yoke, carrying a winding of stout insulated copper wire, the number of turns and consequently the ampere-turns for a given current, being known. A "test piece" of the iron whose magnetic properties are under investigation forms the core of the electro-magnet, the magnetic circuit being completely closed. This necessitates an accurate fit of the test piece in the annular opening at C, and perfect contact of the end of the test piece on the machined surface A.

The test consists of a measurement of the force necessary exactly to neutralise the magnetic attraction between the yoke and the test piece at this surface, this attractive force being caused by a known magneto-motive force due to the winding.

The magnetic force of attraction between two surfaces with an area of contact A cm² is given by the equation :—

$$P = \frac{B^2 A}{8 \pi} \text{ dynes.}$$

(P is used to denote this force, because it is, in common parlance, the "pull" of the magnet.)

If the "pull" is known, B can be obtained by simple manipulation of the formula, thus

$$B = \sqrt{\frac{8 \pi P}{A}}$$

while H , the magnetising force, is equal to $\frac{4 \pi IN}{10l}$.

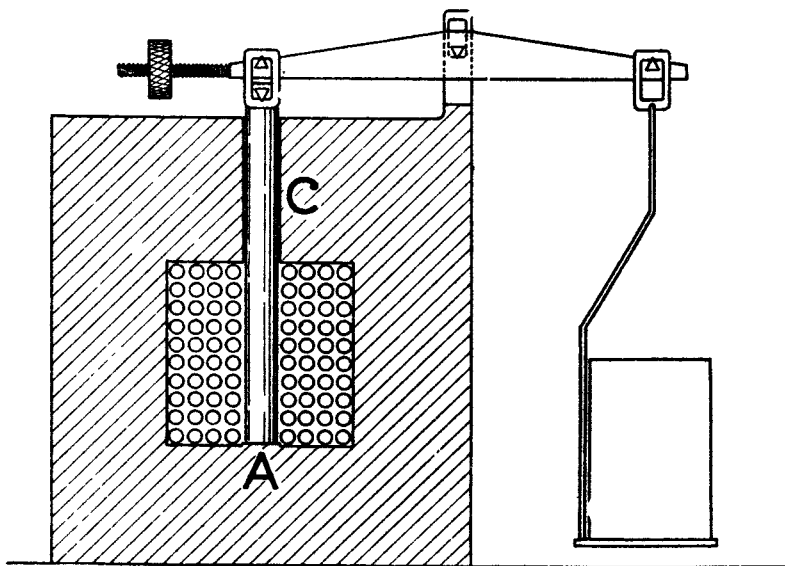


FIG. 15, CHAP. II.—Permeameter.

Example

In the permeameter shewn, the test piece is 20 cms. long and 2 cms. diameter. The magnetising winding carries 200 turns. It is found that a weight of 5,000 grams is just sufficient to overcome the attraction at A, when 1 ampere flows in the coil. Determine the magnetising force, H , the flux density, B , and the permeability of the sample at this flux density.

$$\begin{aligned} H &= \frac{4\pi}{10} \times \text{ampere turns per cm.} \\ &= \frac{4\pi}{10} \times \frac{200}{20} = 4\pi \text{ or } 12.57 \left(\frac{\text{dynes}}{\text{cm}^2} \right) \end{aligned}$$

CHAPTER II.—PARAS. 22-23

This is also numerically equal to the flux density in the gap when the test piece is absent.

$$P = 5,000 \text{ grams or } 4,905,000 \text{ dynes.}$$

$$\text{Now } A = \pi r^2 \text{ and } r = 1 \text{ cm.}$$

$$\therefore A = \pi \text{ cm.}^2$$

$$B = \sqrt{\frac{8 \pi P}{A}} = \sqrt{8 \times 4905000}$$

$$= \sqrt{39,240,000} = 6264 \text{ tubes/cm.}^2 \text{ or gauss.}$$

The permeability is the ratio $\frac{B}{H}$ or $\frac{6264}{12.57}$

$$\therefore \mu = 498.$$

Note.—The reluctance of the massive yoke is neglected in comparison with that of the test piece. A correction for this could be applied if necessary.

22. Typical B/H curves for various ferromagnetic materials are shewn in fig. 16 and the variation of μ with flux density for a particular sample of mild steel, in fig. 17. The molecular

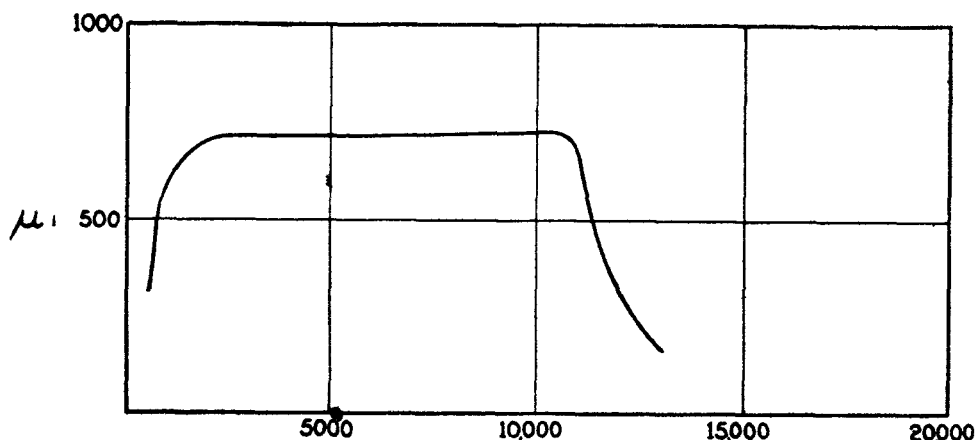
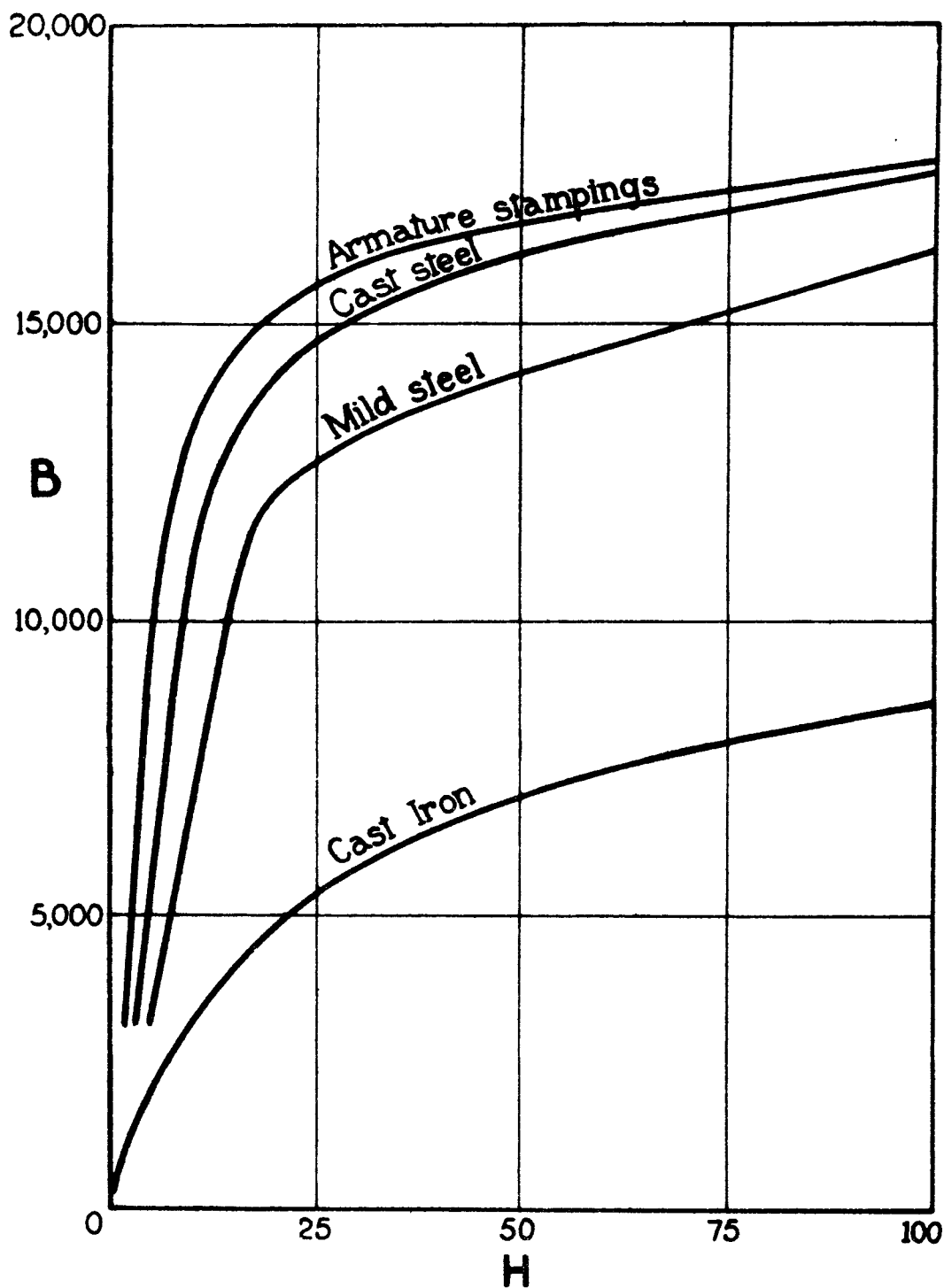


FIG. 17, CHAP. II.— μ/B curve.

theory of magnetism outlined in a preceding section was in part derived from a study of B/H curves of various materials. The variation of flux density, as the magnetising force is increased, is generally somewhat as follows. Commencing with a totally unmagnetised sample, the application of a small magnetising force will result only in the partial disturbance of some of the closed magnetic chains resident in the sample, and does not result in the production of an appreciable external field; the ratio of B to H (i.e. the permeability) is therefore low for small values of B , as shewn in fig. 17. Once the closed magnetic chains have been broken up, the increase of flux density is practically proportional to the increase in magnetising force, and over a considerable range of flux density the permeability is constant. When practically all the molecular magnets have been aligned with their axes in line with the axis of the specimen, further increase of magnetising force is only devoted to a slight improvement of this alignment, little increase in flux density resulting, hence the permeability again falls to a low value. In the latter state the sample is said to be magnetically saturated.

Hysteresis. The hysteresis loop

23. If a series of B/H measurements are taken, with increasing values of H , from zero up to some definite value, say H_1 , and a new series then taken with decreasing values of H , from H_1 to zero, it will be found that the plotted results give two different B/H curves, the latter



B/H CURVES OF IRON AND STEEL

lying above the former, so that on the descending curve when H is zero, B still has a finite value, which is called the remanence of the material. This is an illustration of the fact that any ferromagnetic substance possesses to some extent the property of retentivity, i.e. having been magnetised by some external means, the sample retains its magnetism when the magnetising force has been withdrawn. This remanent or residual magnetism can be removed by applying a magnetising force in such a manner that the sample tends to become magnetised with opposite polarity. The magnetising force thus necessary to overcome the residual magnetism is called the coercive force. If the sample is taken through a complete cycle of magnetisation, i.e. from $H = +H_1$ through $H = 0$, to $H = -H_1$ and then back through $H = 0$ to $H = +H_1$ as shewn in fig. 18, the graphical representation of the B/H relationship is called a hysteresis loop. The magnetisation may be considered to lag behind the magnetising force, and the term hysteresis effect is used to describe the phenomenon. Hysteresis may be regarded as an expression of the work done in overcoming the friction between the molecules of the substance undergoing magnetisation; this work is converted into heat, and is therefore irrecoverable. Hysteresis losses are only of importance in the case of iron subjected to successive cycles of magnetisation.

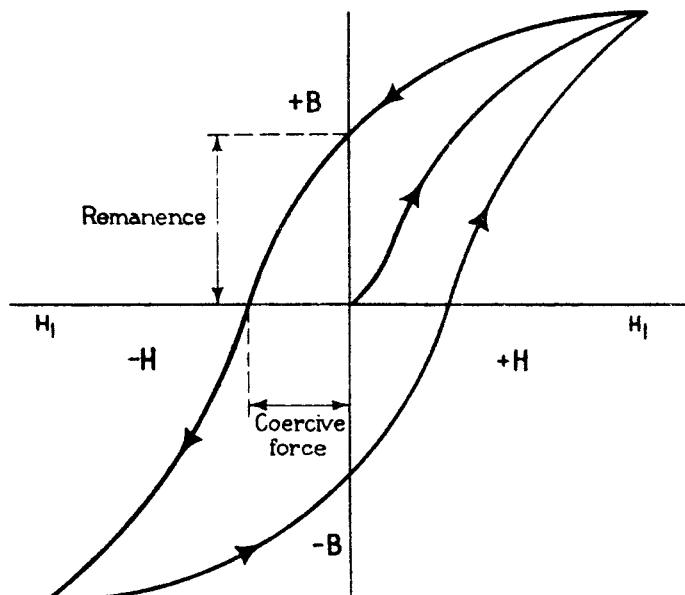


FIG. 18, CHAP. II.—Hysteresis loop.

INDUCED E.M.F.

24. We have seen that whenever there occurs a change in the amount of magnetic flux linking with an electric circuit, the energy expended in changing the flux linkage is partially converted into electrical energy and consequent production of an electromotive force. The flux may be set up by a permanent magnet, an electromagnet, an adjacent current-carrying conductor or a current in the circuit in which the E.M.F. is induced. The methods of producing an induced electromotive force may therefore be classified as follows:—

- (i) Moving flux, stationary conductor, used in rotating field alternators.
- (ii) Stationary flux, moving conductor, used in common forms of D.C. generators and some types of alternator.

CHAPTER II.—PARA. 25

(iii) Varying flux, stationary conductor.

- (a) Mutual induction, a varying flux in one circuit setting up an E.M.F. in an adjacent one.
- (b) Self induction, the variation of flux set up by a change of current in a circuit inducing an E.M.F. in the circuit itself.

A simple example of the first class is depicted in fig. 19 in which a coil of wire is connected to a sensitive galvanometer. When the magnet is stationary, no current flows and there is no deflection of the meter. On dropping the magnet into the coil, the magnetic flux links with each turn of the coil in succession and during the time the flux linkage is changing, the induced E.M.F. causes a flow of current with a resulting deflection of the needle. When the magnet comes to rest (in the position shewn by a dotted outline) the flux is again stationary with regard to the circuit and the induced E.M.F. falls to zero. On withdrawing the magnet the change of flux linkage again produces a momentary E.M.F. but in the opposite direction and this is shewn by a momentary reverse deflection of the galvanometer needle.

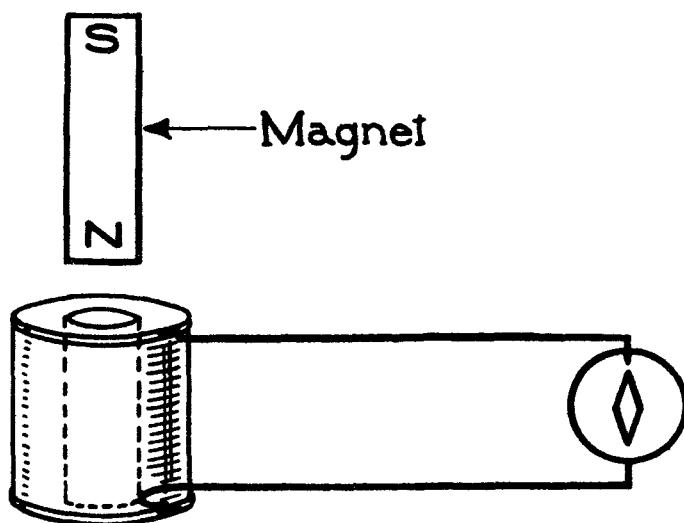


FIG. 19, CHAP. II.—Induction of E.M.F. by motion of magnetic field.

25. An interesting example of the production of electromotive force by the second method is to wind a coil of many turns on a "former" about two feet square, the ends of the winding being connected to a very sensitive galvanometer. Allow the coil to hang vertically by its connecting leads, and turn it sharply through 180° when a deflection of the galvanometer will be observed. The E.M.F. in this case is due to the change of linkage between the coil and the magnetic field of the earth and this apparatus when suitably calibrated can be used to determine the magnetic field strength of the earth at any point.

In order to meet the contingency in which only one portion of a circuit is actually situated in a given field, it is often convenient to speak of the amount of flux cut by a conductor owing to relative motion between conductor and field. As an example, consider a conductor moving across a magnetic field established between two unlike poles, as in fig. 20. The change of flux linkage with the whole circuit is evidently equal to the number of tubes of flux through which the conductor passes, and the conception of cutting is particularly useful because it lends itself to the application of the following mnemonic for finding the direction of the induced electromotive force.

Fleming's right hand rule :—

Extend the thumb, forefinger and middle finger of the right hand in three mutually perpendicular directions. Point the thumb in the direction of Motion, the Forefinger in the direction of the Field, then the Middle Finger gives the direction of the induced E.M.F.

The magnitude of the induced E.M.F. can be derived by application of the formula

$E \text{ (average)} = \frac{N}{10^8} \frac{(\Phi_2 - \Phi_1)}{t}$. If the conductor moves across the field perpendicularly with velocity u centimetres per second and the length of the conductor is l centimetres, the area swept by

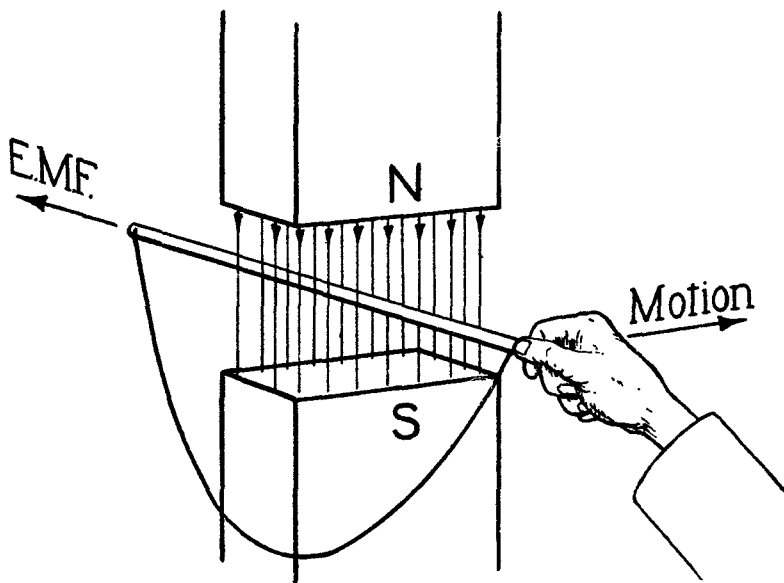


FIG. 20, CHAP. II.—Induction of E.M.F. by motion of conductor.

the conductor per second is lu square centimetres. As $\Phi = BA$, $\Phi_2 - \Phi_1 = B(A_2 - A_1)$ the area $A_2 - A_1$ being the area swept by the conductor, or lu square centimetres per second, hence the change of flux is Blu tubes per second, and is numerically equal to the induced electromotive force in E.M.U. The E.M.F. is thus equal to $Blu \times 10^{-8}$ volts.

If the conductor is not moving through the conductor perpendicularly, but is cutting it at an angle θ , the component of its velocity perpendicular to the flux is $u \sin \theta$ and the induced E.M.F. will be $Blu \sin \theta \times 10^{-8}$ volts. This condition arises in practical dynamo and alternator construction, and further consideration is therefore postponed until Chapter IV.

Faraday's law

26. The phenomenon of electro-magnetic induction was first discovered by Faraday, who summarised the effects in his law of electromagnetic induction, which may now be stated :—

An induced electromotive force is established whenever a change occurs in the magnetic flux linking with an electric circuit. The magnitude of the E.M.F. is proportional to the rate of change of flux linkage.

It must not be supposed that a complete conductive circuit must exist in order that an E.M.F. may be produced. For example, if the circuit consists of a metallic conductor connected to a condenser, the E.M.F. set up by a change of flux linkage in the circuit will set up a conduction current in the metallic portion and a displacement current in the dielectric. Extending this

CHAPTER II.—PARAS. 27-28

principle still further, a change of magnetic flux through a dielectric substance sets up a displacement current in the dielectric, and consequently an electric strain in the material. The electric strain is equivalent to an electric field strength which is measured in volts per metre. This extension of Faraday's law to a dielectric material is the basis of the theory of electromagnetic radiation.

Lenz's law

27. Whenever an induced E.M.F. is set up in an electric circuit and a current thereby established, the conductor experiences a force owing to the interaction between the original flux and that produced by the current in the conductor. The direction in which this force will tend to urge the conductor can be found by applying the left hand rule and is always such as to oppose the motion causing the induced E.M.F. Lenz formulated this principle in his general law :—

Every induced E.M.F. opposes in some manner the change of conditions which produced the E.M.F.

The reader should verify this law by applying the left hand rule to the conductor shewn in fig. 20, in which an E.M.F. is induced by moving the conductor to the right. The force on the conductor due to the flow of current in the circuit will tend to move the conductor to the left.

Mutual induction

28. An electric circuit (A) consisting of a solenoid, battery and switch is shown in fig. 21. On closing the switch a conduction current will be established, and consequently a magnetic flux will thread the coil, its general configuration being shewn in the figure. A portion of the

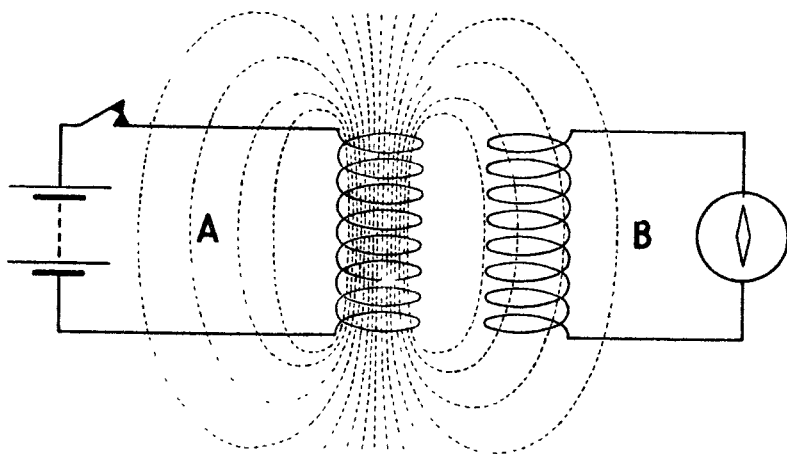


FIG. 21, CHAP. II.—Mutual induction.

flux also links with the adjacent circuit (B). The number of flux linkages evidently depends upon the current in circuit A, the size and shape of the two circuits, and their relative positions. The number of flux linkages common to both circuits is called the mutual flux linkage. It is evident that the mutual flux linkage would be unchanged if the source of E.M.F. were transferred to circuit B and the magnitude of the E.M.F. adjusted so that the current flowing was of the same value as in the original circuit.

Reverting to the arrangement shewn in fig. 21 suppose that a change of current occurs in circuit A ; there will be a change in the total flux produced and consequently a change of flux linkage with circuit B. Now a change of flux linkage, by Faraday's law, gives rise to an E.M.F. and therefore an E.M.F. will be induced in circuit B. It is termed an E.M.F. of mutual induction.

The presence of this E.M.F. can be detected by completing circuit B by means of a sensitive galvanometer, which will indicate a flow of current in circuit B whenever a change of current occurs in circuit A.

Since the flux linkage is proportional to the current, i , and to a constant depending upon the relative shapes and sizes of the circuits, we may write

$$\text{Flux linkage} = M i$$

where M is the constant referred to above. It is called the mutual inductance between the two circuits.

The induced E.M.F. is numerically equal to the time rate of change of flux linkage, that is

$$E = \text{rate of change of } M i$$

Since M is by definition constant

$$E = M \times \text{rate of change of } i$$

which is frequently written

$$E = M \frac{di}{dt} \quad \dots \dots \dots (a)$$

The symbol $\frac{d}{dt}$ will be used frequently in subsequent paragraphs as an abbreviation for "the rate of change with respect to time, of.....". Thus $\frac{d}{dt}i$, or $\frac{di}{dt}$ must be thought of as denoting "the rate of change of current with respect to time".

The equation given above serves to define a unit of mutual inductance. In electromagnetic units the mutual inductance of a circuit is one unit if E.M. unit voltage is induced in it when the rate of change of current is one E.M. unit per second. Similarly in practical units the mutual inductance is one unit if one volt is induced in it when the rate of change of current is one ampere per second. This practical unit of mutual inductance is called the Henry. It is equal to 10^9 E.M. Units.

Calculation of mutual inductance

29. The mutual inductance between two circuits is not readily calculated in any but the simplest instances. A case which lends itself to calculation is that of two toroidal coils on the same former, one winding being wound over the other. It may then be assumed that when a current is set up in either coil the whole of the magnetic flux will link with both coils, or (introducing a term frequently employed to convey the same meaning) no magnetic leakage occurs. The permeability of the material will also be taken as constant. Let a varying current of intensity i amperes at any moment flow in one coil which has N_1 turns. Then the flux set up is given by the equation

$$\Phi = \frac{4\pi}{10} \frac{iN_1}{S},$$

S being the reluctance of the magnetic path. Now only a change of flux can produce an E.M.F. in the second circuit and since every factor of the right hand member of the equation is constant except the current, the rate of change of flux with respect to time, $\frac{d\Phi}{dt}$, must be

$$\frac{d\Phi}{dt} = \frac{4\pi N_1}{10 S} \frac{di}{dt}$$

CHAPTER II.—PARAS. 30-31

and this is also equal to the E.M.F. induced in each turn of the winding N_2 . The total E.M.F. in this winding is therefore

$$e = \frac{4\pi}{10} \frac{N_1 N_2}{S} \frac{di}{dt} \text{ E.M.U.}$$

$$\text{or } E = \frac{4\pi}{10^9} \frac{N_1 N_2}{S} \frac{di}{dt} \text{ volts} \quad \dots \dots \dots (b)$$

comparison of this equation with equation (a) shews that both can only be true if M (in henries) is equal to $\frac{4\pi}{10^9} \frac{N_1 N_2}{S}$.

Self-induction—Inductance

30. Referring to fig. 21 above, let it be supposed that the circuit B is non-existent. On closing the switch, a conduction current and consequently a magnetic flux is established just as before. Now the magnetic flux links with the circuit A, and any change of current will result in a change of flux linkages through the circuit. The flux linkages are proportional to the intensity of the current and also to a constant L , which depends upon the shape and size of the coil. It is therefore permissible to write

$$\text{Flux linkage} = L i$$

The constant L is called the coefficient of self-induction, or the inductance of the circuit, the latter expression being usual in radio practice. Now as the value of the induced E.M.F. in any circuit is numerically equal to the rate of change of flux linkages,

$$E = \frac{d}{dt} (L i)$$

since L is constant by definition

$$E = L \frac{di}{dt}$$

The unit of inductance is defined in exactly the same way as the unit of mutual inductance. The practical unit is the Henry, and is the inductance of a circuit in which a current change of one ampere per second sets up an induced E.M.F. of one volt.

By Lenz's law this self-induced E.M.F. must oppose the change of current which produced it, that is, it is actually proportional to $-\frac{di}{dt}$. This tendency of the induced E.M.F. may therefore be indicated by writing

$$E = -L \frac{di}{dt} \quad \dots \dots \dots (c)$$

Calculation of inductance

31. Referring to the example of calculation of mutual inductance given above, the flux set up by the instantaneous current of i amperes is

$$\Phi = \frac{4\pi}{10} \frac{iN}{S}$$

the rate of change of the flux is

$$\frac{d\Phi}{dt} = \frac{4\pi}{10} \frac{N_1}{S} \frac{di}{dt}$$

and the total change of flux linkage with the N_1 turns of the winding will therefore be $\frac{4\pi}{10} \frac{N_1^2}{S} \frac{di}{dt}$, which is equal to the induced E.M.F. in E.M.U.

In practical units

$$E = -\frac{4\pi}{10^9} \frac{N_1^2}{S} \frac{di}{dt} \quad \dots \dots \dots (d)$$

the minus sign being inserted in order to satisfy Lenz's Law. Comparison with equation (c) above indicates that

$$L = \frac{4\pi}{10^9} \frac{N_1^2}{S} \text{ henries}$$

This formula would be perfectly general if two conditions were always satisfied, that is, if it were certain that the whole of the flux linked with every turn of the winding, and if the reluctance S could be calculated for any shape or size of circuit. In practice, the formula is used in conjunction with a form factor which takes the "geometry" of the circuit into consideration. (The term geometry is nowadays frequently used as an omnibus term for "shape, size and disposition of parts").

The henry is too large for general use, although coils of 1,000 henries inductance are occasionally met with, e.g. the secondary winding of a high class intervalve transformer. For use in radio practice the henry is subdivided as under

$$\begin{aligned} 1 \text{ millihenry} &= \frac{1}{1,000} \text{ henry, or } 10^{-3} \text{ henry} \\ 1 \text{ microhenry} &= \frac{1}{1,000,000} \text{ henry, or } 10^{-6} \text{ henry} \\ 1 \text{ absolute unit} &= \frac{1}{1,000,000,000} \text{ henry, or } 10^{-9} \text{ henry.} \end{aligned}$$

The absolute unit of inductance is also called the centimetre. This can be justified on theoretical grounds, and practically by the fact that the inductance depends upon the size and shape of the circuit, that is upon measurements of length.

Effect of inductance in an electrical circuit

32. When an electrical circuit possesses the property of inductance, and it must be remembered that no circuit can be entirely without it, the effect of its presence is to oppose any change in the value of the current, or alternatively any alteration in the state of the electrons or electric charges from a state either of rest or of uniform motion. Now in connection with matter, the property having the same nature is called inertia and we may therefore say the inductance is the electrical analogue of inertia. To take a concrete example, consider the circuit shewn in fig. 22, in which the battery of E.M.F. E volts is connected to the coil of inductance L henries and

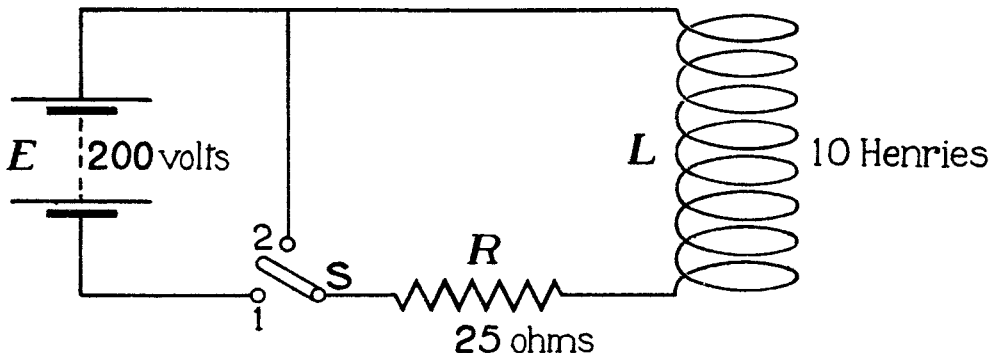


FIG. 22, CHAP. II.—Circuit possessing resistance and inductance.

CHAPTER II.—PARA. 33

R ohms resistance by means of a switch S , which is so designed as to short circuit the coil at the instant of disconnecting the battery. On placing the switch S into position 1, a current will be established. Now by Ohm's law, this current is equal to $E \div R$, without reference to time, and if the circuit were absolutely non-inductive the current would assume this value instantaneously.

If the applied E.M.F. is 200 volts, $R = 25$ ohms and $L = \text{zero}$, the current would be 8 amperes from the instant of closing the switch until the circuit was broken, when the current would fall to zero instantly. On the other hand, if the circuit possessed an inductance L of 10 henries, but offered absolutely no resistance, the current would commence to grow uniformly at a rate of $\frac{E}{L}$ or 20 amperes per second, and would ultimately reach an infinite value. If we assume for purely theoretical purposes that the current cannot exceed the value $\frac{E}{R}$ to which it would be limited by the presence of resistance, the current, growing at the rate of $\frac{E}{L}$ amperes per second, would reach this value in a time T which is given by the relation

$$\frac{ET}{L} = \frac{E}{R}$$

$$T = \frac{L}{R}$$

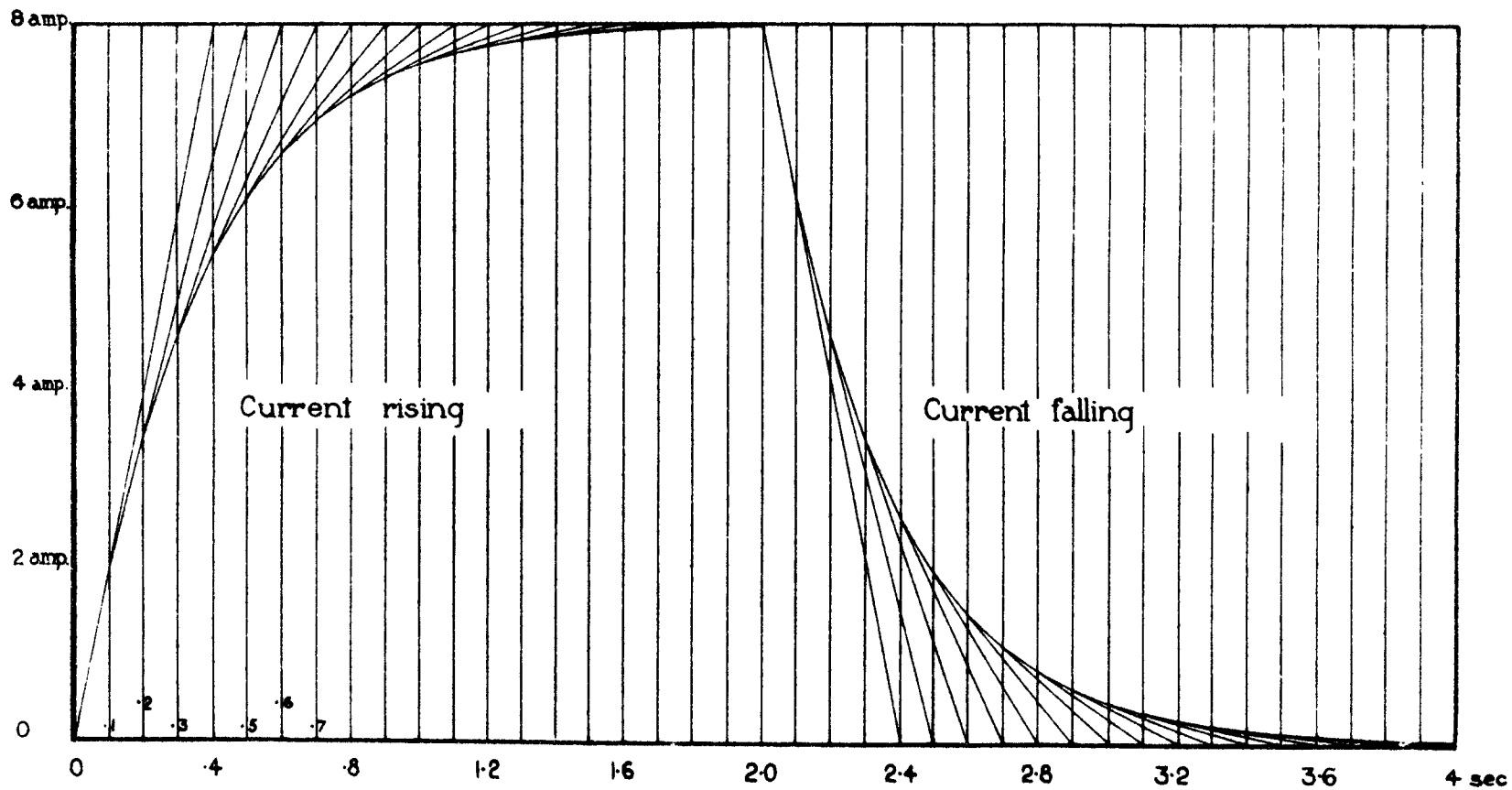
With the circuit constants assumed above, $T = \frac{10}{25} = .4$ second. The time taken by the current to reach the value $\frac{E}{R}$, assuming that its original rate of increase were maintained, is called the time constant of the circuit. It is analogous to the time constant of the circuit possessing both capacitance and resistance, the charging and discharging processes of which were explained in Chapter I. The initial rate of increase of current cannot be maintained, however, because the growth of current through the coil sets up around it a magnetic flux of increasing density, which links with the conductor and induces in it an E.M.F. (Faraday's law). In accordance with Lenz's law this E.M.F. tends to oppose the change of conditions which produced it, i.e. it opposes the growth of current, and is said to be a counter-E.M.F.

33. After a short interval of time, say .1 second, the current has risen to 2 amperes, the P.D. across the resistance will be $2 \times 25 = 50$ volts, and the voltage available to overcome the electrical inertia of the inductance, or to increase the magnetic flux, is only $200 - 50 = 150$ volts.

The current will continue to rise, but at a rate of $\frac{150}{L}$ or 15 amperes per second instead of 20 amperes per second as originally, and if this rate were maintained the current would reach 8 amperes in a further .4 second. The growth of current can in fact be obtained by a graphical construction identical with that used in Chapter I for the charging of a condenser, and fig. 23 shows the current growth with the circuit constants given above. It will be observed that by

this graphical construction, the current reaches 68.5 per cent. of its final value in the time $T = \frac{L}{R}$.

If sufficiently small time intervals are taken, however, a more accurate curve results, and in reality the current would reach 63.2 per cent. of its final value in the time T . The graph shows that the current would not reach the value 8 amperes given by Ohm's law until about two seconds had elapsed from the time of completing the circuit. It must not be thought that this phenomenon invalidates Ohm's law; it really proves it, for if we say that $I = \frac{E}{R}$, in this particular instance, a major error is committed, because the total E.M.F. acting in the circuit



RISE AND FALL OF CURRENT IN AN INDUCTIVE CIRCUIT

is not E , but $E - L \frac{di}{dt}$, so that the true relation between current, E.M.F. and the circuit constants is

$$i = \frac{E - L \frac{di}{dt}}{R}$$

$$\text{or } Ri + L \frac{di}{dt} = E$$

the small letter i being used to indicate that this equation gives the instantaneous value of the current at some interval of time after closing the switch. The mathematical solution of this equation is given in a note at the end of the chapter.

34. Reverting to the circuit diagram of fig 22, suppose that after the current has been established for some time the switch is suddenly placed in position 2. The inductance and resistance are then disconnected from the supply and short-circuited upon themselves. The current through the inductance now commences to die away, and the decreasing flux, which collapses into the conductor, sets up an induced E.M.F. which opposes the change of current, and therefore tends to maintain it at its original value. If there were no resistance the current would commence to fall at a rate of $\frac{E}{L}$ amperes per second, as shown graphically in the figure. In the first 0.1 second the current would fall to 6 amperes, and the P.D. across the resistance would be 150 volts. The current then continues to fall at a rate of $\frac{150}{L}$ or 15.0 amperes per second, and its rate of decrease becomes less and less as time goes on. The "decay" curve, as it is called, is obtained by the graphical construction previously outlined.

If instead of short-circuiting the coil by means of the switch the circuit is simply broken, the effect of the voltage gradient (i.e. the electric field strength in volts per centimetre, vide Chapter I) across a minute gap of air at the instant of metallic disconnection, causes ionisation of the air in the gap which then becomes partially conductive, and the current continues to flow across the gap in the form of an electric arc. This arcing is accentuated by the presence in the circuit of coils of large inductance, and in such circuits some steps are usually taken to reduce such effects of the arc as burning of the switch contacts, one method adopted being to connect a condenser in parallel with them.

Energy stored in a magnetic field

35. Just as energy is stored in the electric field when an electrical condenser of capacity C farads is charged to a difference of potential of V volts, the energy stored being $\frac{1}{2} CV^2$ joules, so when a magnetic field is established it can be regarded as a storage of energy. While the magnetic field is being established the current grows slowly to a final steady value of $\frac{E}{R}$, E being the E.M.F. in the circuit and R its total resistance. The amount of energy stored can be found as follows:—

The current starting from zero value, reaches the final value $I = \frac{E}{R}$ in t seconds, the average rate of change of current being therefore $\frac{I}{t}$ amperes per second, and the average current $\frac{L}{2}$ amperes. The average counter-E.M.F. by Faraday's law is $L \times \text{rate of change of current}$ or $\frac{LI}{t}$ volts, L being in Henries.

The average rate at which energy is stored is therefore $\frac{LI}{t} \times \frac{I}{2}$ volt-amperes, which is the power expended in creating the magnetic field. Since this power is only expended during the time the current is growing, the expenditure goes on for t seconds, and the work done is $\frac{LI}{t} \times \frac{I}{2} \times t$ or $\frac{1}{2} LI^2$ joules.

CHAPTER II.—PARAS. 36-37

It has been said that the energy is stored, the implication being that it is restored to the source of supply at some future period. This takes place when the current is caused to fall to zero, for example by withdrawing the original source of energy from the circuit. The collapse of the magnetic field results in a change of flux linkage with the circuit and an E.M.F. is set up which by Lenz's law tends to oppose the fall of the current. As a result the current does not immediately fall to zero when the E.M.F. is withdrawn but dies away slowly and eventually becomes zero when all the stored energy has been expended.

Energy density

36. In Chapter VII it is shown that under certain conditions an electromagnetic wave may be radiated from an electric circuit, this wave consisting of an electric field such as is discussed in Chapter I, and a magnetic field of the nature just considered. In the theoretical consideration of this radiation it is convenient to refer to the amount of energy stored in magnetic form per unit volume of space, which is called the energy density. The energy density of the field inside the winding of a toroidal coil can easily be found if it is assumed that the flux density over the whole cross-section of the coil is uniform. Let the toroid carry N turns, the length of the mean magnetic line passing through all the turns be l centimetres, and the cross-section of the coil be

A square centimetres. Then the inductance of the coil is $\frac{4 \pi N^2 A \mu}{l}$ E.M. units, and the energy

stored in the magnetic field is $\frac{1}{2} L I^2$ ergs. if both L and I are expressed in E.M.U. Hence

$$W = \frac{1}{2} \times \frac{4 \pi N^2 A \mu}{l} I^2 \text{ ergs.}$$

But as the volume of the whole magnetic field inside the coil is $A l$ cubic centimetres, the energy stored in each cubic centimetre is

$$\frac{4 \pi I^2 N^2 A \mu}{2 l} \times \frac{1}{A l} = \frac{4 \pi I^2 N^2 \mu}{2 l^2} \text{ ergs.}$$

and the energy density is

$$\left(\frac{4 \pi I N}{l} \right)^2 \times \frac{\mu}{8 \pi} \text{ ergs per cm.}^3$$

As $\frac{4 \pi I N}{l}$ is the magnetic field strength inside the coil, it may be denoted by H and therefore

the energy density is $\frac{\mu H^2}{8 \pi}$ ergs per cubic centimetre. This result is true for any distribution of the magnetic field.

TYPES OF INDUCTANCE

37. A coil which has been deliberately produced for the purpose of introducing the property of inductance into a circuit is properly called an inductive coil or inductor, but it is more generally termed simply an inductance. Inductances may be classified as large or small, the former term indicating those having a value of the order of henries, and these are constructed by winding many turns of wire upon a closed iron core, or upon a similar core having a small air gap. Such coils are only employed when the frequency is comparatively low. Small inductances are those having an inductance of the order of microhenries, and are either without iron cores, or have cores of a special iron alloy composition. A coil wound upon a core having unit permeability is always spoken of as an "air core" inductance, irrespective of the material upon which it is actually wound, because the latter has no influence upon the value of the inductance, which depends solely upon the geometry of the winding.

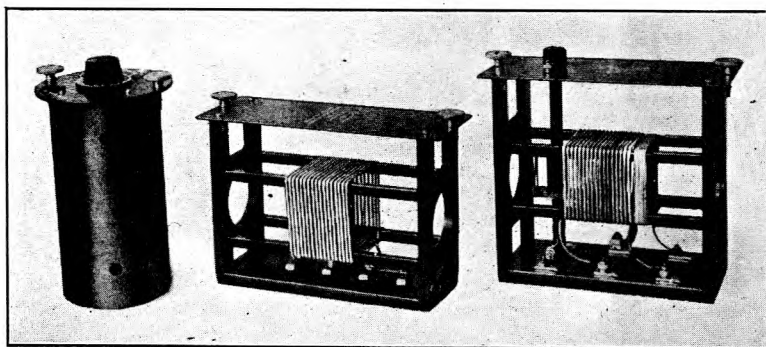


FIG. 24A, CHAP. II.—Typical air-core inductances.

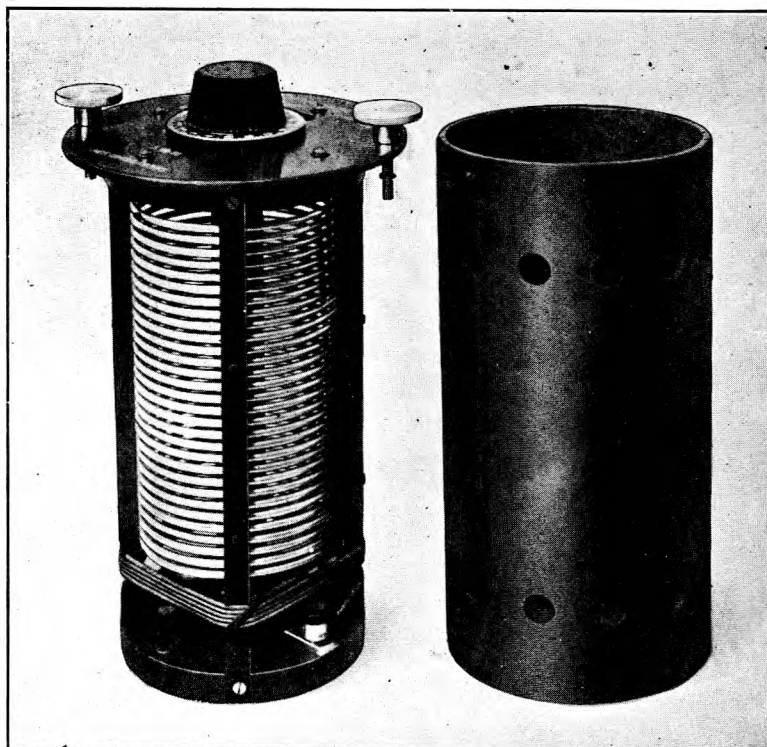


FIG. 24B, CHAP. II.—Continuously variable air-core inductance.

38. Air core inductances are met with in many different sizes and shapes as components of radio transmitters and receivers. Three typical coils are shown in fig. 24a. The construction of the one on the left is shown in fig. 24b. The coil is of copper tubing, silver plated in order to avoid oxydisation, and so reduce the surface resistance. The inductance is continuously variable within certain limits by means of a spring clip which embraces the conductor. The clip is carried on a radial arm which is mounted upon a nut working on a fixed brass screw extending along the whole length of the axis of the coil. The pitch of the thread is the same as that of the winding, and the screw is surrounded by a guide tube or sleeve of insulating material which is slotted along its length to allow the nut, and consequently the radial arm, to move axially. The insulating sleeve is rotated by means of the milled head shown, and as the screw cannot rotate the spring contact arm is constrained to travel round the turns of the coil. An indicating device is fitted in order to show the position of the contact.

The two inductances on the right (fig. 24a) are of fixed value, the winding being of stranded wire wound upon an insulating framework. That on the extreme right actually carries three distinct coils, the main winding being the stoutest conductor; the turns are spaced by a distance approximately equal to the overall diameter of the conductor. On the right a second winding is carried on the same framework and is linked with the main winding by mutual induction. The third winding is carried on a smaller framework which is mounted upon a vertical bar. This winding is also linked with the main coil by mutual induction, the flux linkage being variable by rotating the smaller coil by means of a knob on the upper panel.

Receiving inductances are usually of solid wire although multistranded wire is sometimes used. When the inductance is to be adjustable in large steps, a series of tappings is made and the points connected to a rotary switch. When the value of inductance is to be adjustable to fine limits the variometer construction is adopted.

The variometer

39. The variometer principle depends upon the presence of mutual inductance between two circuits. If two coils of inductance L_1 and L_2 respectively are separated by such a distance that the interlinkage of their magnetic fields is negligible, the inductance of the two in series can be shown to be $L_1 + L_2$ (Chapter V). If, however, they are brought closely together so that the field of each coil interlinks with the turns of the other, mutual inductance exists between them. An applied E.M.F. E will then cause a current i to flow through both coils and the rate of increase of the current must be the same in each. If the coils and their respective inductances are designated by L_1 and L_2 , and the mutual inductance between them by M , then owing to the rise

of current there is a counter-E.M.F. in L_1 , due to its self induction, its value being $-L_1 \frac{di}{dt}$

and as the growing field of L_2 is also threading L_1 , an additional counter-E.M.F. $-M \frac{di}{dt}$.

In the second coil, the growth of its own flux causes a counter-E.M.F. $L_2 \frac{di}{dt}$, and the growth of the

flux of L_1 , which also embraces L_2 , causes an additional counter-E.M.F. in it which is equal to

$-M \frac{di}{dt}$. Hence the total counter E.M.F. is

$$-(L_1 + L_2 + 2M) \frac{di}{dt}.$$

But as the perfectly general expression for the counter-E.M.F. in such conditions is $-L \frac{di}{dt}$,

L being the total inductance of the circuit, the latter is equal to $L_1 + L_2 + 2M$. It will be noted that in the above example, the mutual flux linkage is so disposed that the counter-E.M.F. due to it is in the same direction as the counter-E.M.F. of self induction. This is not necessarily so, and

CHAPTER II.—PARA. 39

in fact the counter-E.M.F. of mutual induction in both coils will be reversed if the direction of the field of one coil is reversed. This can be achieved either by actually turning the coil through 180° or by reversing the connections to its ends. The above reasoning again applied then shows that the total inductance is given by

$$L = L_1 + L_2 - 2M.$$

The fact that mutual inductance may have either positive or negative sign, while self-inductance is always positive, is of considerable importance. An immediate application is the variometer (see fig. 25). In this instrument two coils are mounted concentrically with each other, and connected in series. The inner coil is arranged to rotate with reference to the outer through an angle of 180° , and the mutual inductance between the two coils is therefore variable from $+M$ to $-M$.

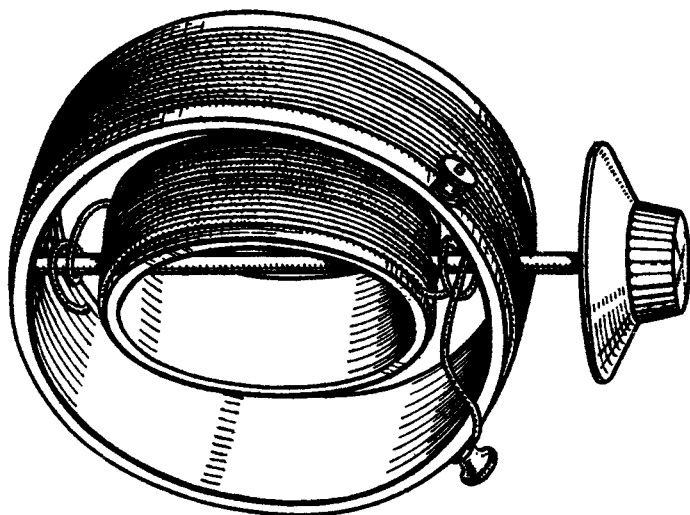


FIG. 25, CHAP. II.—Variometer.

The whole inductance of the two coils in series including this mutual inductance is from $L_1 + L_2 - 2M$ to $L_1 + L_2 + 2M$, and the variation is perfectly smooth throughout this range. Instead of being connected in series, the two coils may be connected in parallel. The value of the inductance is then rather more difficult to deduce and will be given in Chapter V. It is shown in that chapter that the inductance of two coils in parallel without mutual inductance between them

is

$$L = \frac{L_1 L_2}{L_1 + L_2}$$

while if a mutual inductance M exists between them

$$L = \frac{L_1 L_2 - M^2}{L_1 + L_2 \pm 2M}.$$

It should hardly be necessary to add that the same units of inductance must be used throughout; if L_1 is given in μH , L_2 and M must also be expressed in μH , while if L_1 is given in henries the same unit must be adopted for the other values before insertion into the equation.

Where space is limited it may be found desirable to utilise the basket-weave coil. This is wound upon a disc, which has an odd number of radial slots, generally seven or nine. These slots extend from the circumference inwards to a depth of about one half of the overall radius. The winding is commenced at the inner end of one slot, passing the wire alternately to one side of the disc or the other, as shown in fig. 26. The inductance of such a coil may be calculated by

formula (b) below. Many formulae have been developed for the calculation of the inductance of multilayer coils of various shapes, but it is frequently more practicable to measure the inductance by a practical method than to calculate it. An example of such a measurement is given in the chapter dealing with radio-frequency measurements.

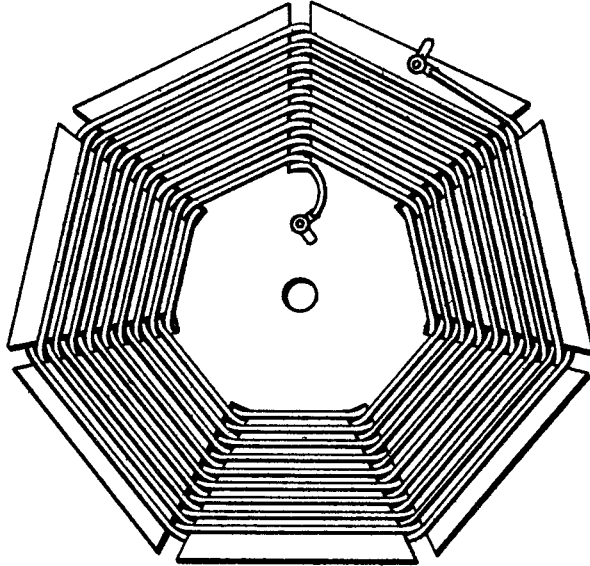


FIG. 26, CHAP. II.—Basket weave coil.

Calculation of inductance of air-core coils

40. In certain emergencies it may be necessary to construct extempore inductance coils for some special purpose, and in these circumstances it is usually convenient to adopt the single layer solenoidal method of winding. The form factor for a single layer solenoid is a complex mathematical function of the ratio of length to diameter of winding, and a formula often used is

$L = \frac{d}{2} N^2 f$ where d is the diameter of the coil measured from centre to centre of the wire, N is

the number of turns and f is the form factor. For values of $\frac{\text{length}}{\text{diameter}}$ lying between .1 and 10, the

latter can be represented approximately by an empirical formula $f = \frac{r}{8.5r + 10l}$, and the above expression for inductance then becomes,

$$L = \frac{r^2 N^2}{8.5r + 10l} \quad \dots \quad (a)$$

where L is the inductance in microhenries, r the mean radius and l the length of the solenoid, all measurements being in inches. The former upon which the coil is wound is assumed to be of circular section, but if a hexagonal or octagonal former is used the effective radius may be taken the mean of the radii of the inscribed and circumscribed circles on the section. If the coil is of solenoidal form, but having more than one layer, a similar formula may be used, but the radius must be half the mean diameter of the winding, that is

$$r' = \frac{d_1 + d_2}{4}$$

CHAPTER II.—PARAS. 41-42

d_1 being the diameter of the former and d_2 the diameter of the wound solenoid, while for the length, the quantity $l' = l + \frac{1}{2}(d_2 - d_1)$ must be substituted. The inductance can then be calculated from the formula

$$L = \frac{(r')^2 N^2}{8.5 r' + 10 l'}$$

THE TELEPHONE RECEIVER

41. The first instrument handled by the embryo wireless operator is generally the telephone receiver, or rather the head set comprising the headband, cord and a pair of receivers. The telephone receiver is a device which converts variations of electrical current into sound waves, that is, electrical energy into mechanical energy. It consists of a permanent magnet upon which are mounted two soft iron pole pieces, each carrying a magnetising winding. The two coils are connected in series in such a manner that current in a given direction strengthens both poles, while current in the opposite direction weakens them. A circular diaphragm of soft iron or stalloy is so mounted that both pole pieces normally exert a slight pull on its centre, although the diaphragm does not quite touch the pole pieces. The general arrangement is shewn in fig. 27.

42. The action of the telephone receiver may be explained with reference to the circuit given in fig. 20 Chapter I in which it is placed in series with a carbon microphone and a suitable

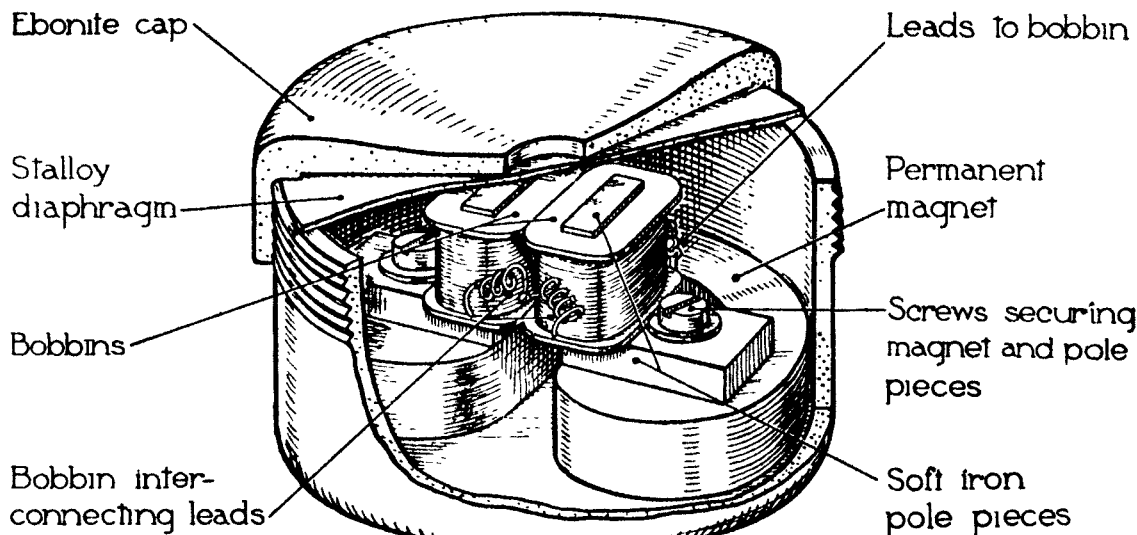


FIG. 27, CHAP. II.—Telephone Receiver.

battery. When the microphone is unaffected by sound (a term frequently used to express this state being "quiescent") a small steady current flows round the coils of the telephone receiver, which sets up a magnetic flux in addition to that provided by the permanent magnet. The direction of current should be such that the two fluxes are additive, and the diaphragm will be attracted rather more closely to the pole pieces than in the absence of this polarising current. On speaking into the microphone, however, the changes in the resistance cause changes in the value of the current round the windings, and corresponding changes in the magnetic flux. An increase of flux will cause further attraction of the diaphragm, and a decrease will result in its partial release from attraction, so that the diaphragm is set into vibration in a manner corresponding to the variation of current, and a sound is emitted by the diaphragm which

resembles that originally impressed upon the microphone. The necessity for the inclusion of the permanent magnet is not obvious from the foregoing explanation, for if it were omitted, the variation of current in the winding round the soft iron pole pieces would still cause a varying pull on the diaphragm. The permanent magnet has two functions (i) it gives the receiver greatly increased sensitivity for the same current changes; (ii) in its absence the vibration of the telephone diaphragm would take place at twice the rate of vibration of the microphone diaphragm, and the sound emitted by the telephone would not closely resemble the original sound. Taking these points in order, it will first be shown that the effect of the permanent magnet is to give a greater "pull" on the diaphragm for a given variation of current in the winding than would be obtained in its absence. Let the flux density in the air gap between pole pieces and diaphragms be B . The pull on the diaphragm is given by the equation:—

$$P = \frac{B^2}{8\pi} \text{ dynes per square centimetre (para. 21).}$$

The flux density B may be separated into two components, viz., B_p caused by the permanent magnet and the steady component of current, and $\pm B_c$ caused by the variation of current. The above equation then becomes

$$P = \frac{(B_p \pm B_c)^2}{8\pi} \frac{\text{dynes}}{\text{cm.}^2}$$

or omitting the divisor 8π which will not affect the argument

$$P = (B_p \pm B_c)^2$$

or

$$P = B_p^2 \pm 2B_pB_c + B_c^2.$$

The pull P_0 on the diaphragm in the quiescent condition can be obtained by observing that the component B_c is then zero, hence

$$P_0 = B_p^2$$

and the additional pull due to the variation of current is

$$P_c = \pm 2B_pB_c + B_c^2.$$

This is obviously the portion of the attractive force with which we are immediately concerned. If the two components of flux density B_p and B_c are acting in the same direction, the pull on the diaphragm will be found by using the positive sign in the above expression. If the two fluxes are in opposition, the negative sign will be appropriate.

Examination of this equation shows that if the flux density B_p is zero, that is if no permanent magnet (or its electrical equivalent) is fitted, the attractive force will be $P_c = B_c^2$ only. The presence of the permanent magnet increases the pull caused by a given current from B_c^2 to

$\pm 2B_pB_c + B_c^2$, that is, in the ratio of 1 to $1 + 2\frac{B_p}{B_c}$. As B_p may be hundreds of times as

large as B_c it is apparent that the sensitivity of the receiver is increased enormously by the presence of the permanent magnet. It was stated above that if the current circulated round the pole pieces in such a direction that the resulting flux was of the same polarity as that of the permanent magnet the two fluxes were additive. This is a desirable state of affairs, giving the equivalent of a permanent magnet of even greater pole strength than that of the actual magnet. An additional reason for so arranging the direction of current is that unless this uniformity of polarity is maintained, the two fluxes will be in opposition and the magnetising force of the current will tend to demagnetise the permanent magnet and the sensitivity of the telephones will be impaired. It is therefore desirable to check the polarity of the connections of telephone receivers which are intended for use in the above circumstances and always connect them in such a way that the direct current will set up a flux tending to strengthen the pole pieces.

In order to avoid this necessity it is usual in modern practice to connect the microphone and battery to one winding of a small transformer, forming what is called the primary circuit, while the telephones are connected to the ends of a secondary winding. Under these conditions no steady component of current circulates round the telephone windings, but any variation of

CHAPTER II.—PARA. 43

current in the primary winding causes a variation of flux which embraces the secondary winding and induces in the latter an E.M.F. which at all instants varies in magnitude in accordance with the current variations. As the secondary circuit is completed by the telephone windings, a varying current will be established and will affect the diaphragm in the manner previously explained.

43. Attention may now be devoted to the second reason given above for the inclusion of a permanent magnet. Various possible conditions are illustrated in fig. 28. The first series of diagrams shews the state of affairs when the permanent magnet is absent, but with a large D.C. component of current in the windings, the latter serving the same purpose as the permanent magnet to some extent. It will be observed that the movement of the diaphragm, and therefore the sound emitted from the receiver, is a copy of the original sound wave.

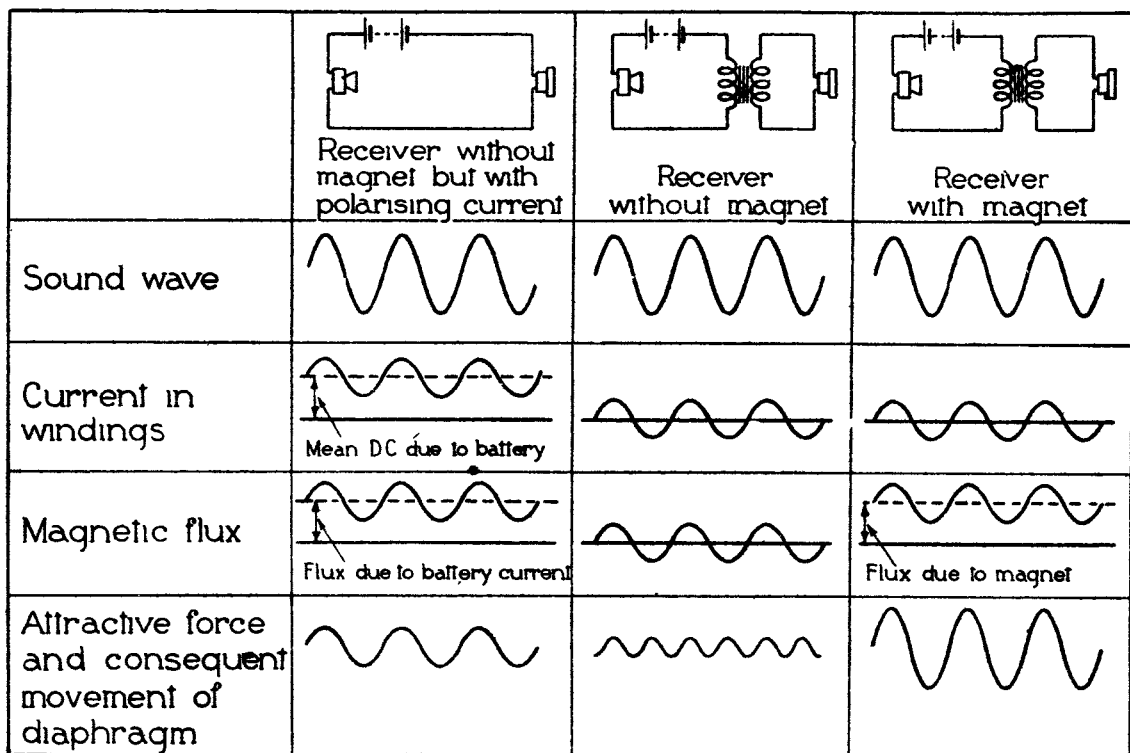


FIG. 28, CHAP. II.—Effect of polarising flux in telephone receiver.

In the second and third series of diagrams the circuit contains a transformer as dealt with above. If no permanent magnet is fitted, the diaphragm is initially under no strain whatever, but will be attracted whenever a current is established in the winding, no matter what its direction may be. The result of this is that the diaphragm is set into vibration in such a way that a single variation of current from zero to some peak value, back to zero, then to an equal peak value in the opposite direction, and finally to zero, causes two separate pulls on the diaphragm, and the latter vibrates at twice the rate at which the current variations are taking place, hence the sound emitted by the receiver is not a copy of the original sound wave. The effect of the permanent magnet is shown in the third series. Here the variation of flux caused by the current variation is superimposed upon the steady flux, and an increase of current causes an increased pull, while when the current decreases the reverse is the case, hence the diaphragm undergoes one variation of displacement for each complete change of current.

NOTE.—Solution of the equation $Ri + L\frac{di}{dt} = E$.

The equation must first be rearranged thus,

$$\frac{di}{dt} + \frac{R}{L}i = \frac{E}{L} \quad \dots \dots \dots 1.$$

Multiplying each term by $e^{\frac{R}{L}t}$

$$\frac{di}{dt} e^{\frac{R}{L}t} + \frac{R}{L}i e^{\frac{R}{L}t} = \frac{E}{L} e^{\frac{R}{L}t} \quad \dots \dots \dots 2.$$

The left-hand side is equal to $\frac{d}{dt} (i e^{\frac{R}{L}t})$

where both i and t are variables, for

$$\frac{d}{dt} (i e^{\frac{R}{L}t}) = \frac{R}{L}i e^{\frac{R}{L}t} + e^{\frac{R}{L}t} \frac{di}{dt}$$

Hence, integrating both members of the equation

$$\begin{aligned} i e^{\frac{R}{L}t} &= \frac{E}{L} \int e^{\frac{R}{L}t} . dt \\ &= \frac{E}{R} e^{\frac{R}{L}t} + C \quad \dots \dots \dots 3. \end{aligned}$$

where C is a constant depending upon the initial conditions.

$$\text{Finally,} \quad i = \frac{E}{R} + C e^{-\frac{R}{L}t} \quad \dots \dots \dots 4.$$

To determine C , we observe that $i = 0$ when $t = 0$, and that both E and R are constant. Hence, at the instant $t = 0$,

$$0 = \frac{E}{R} + C e^0$$

$$\text{or } C = -\frac{E}{R}, \text{ whence}$$

$$i = \frac{E}{R} (1 - e^{-\frac{R}{L}t}) \quad \dots \dots \dots 5a.$$

t being restricted to positive values only, for a negative value of t means a time previous to the closure of the circuit, and it must be assumed that the current is zero until this instant. The equation shewing the decay of the current can be derived from the preceding by varying the initial conditions. If at the time $t = 0$, E becomes zero, equation (4) becomes

$$i = C e^{-\frac{R}{L}t}.$$

CHAPTER II.—PARA. 43

and to determine C , we observe that at the time $t = 0$, i has some value which may be denoted by I_0 . If the current has been flowing for a period sufficiently long, the value of I_0 is given by

Ohm's law, i.e. $I_0 = \frac{E}{R}$

and
$$i = \frac{E}{R} \epsilon^{-\frac{R}{L}t} \quad 5b.$$

The equations (5a) and (5b) are frequently referred to as the equations of logarithmic growth and decay.

It will be observed that when $t = \frac{L}{R}$ these equations become

$$i = \frac{E}{R} (1 - \epsilon^{-1}) \quad \dots \quad 6a \text{ (current growing)}$$

$$i = \frac{E}{R} \epsilon^{-1} \quad \dots \quad 6b \text{ (current decaying).}$$

Now $\epsilon = 2.71828$, $\epsilon^{-1} = \frac{1}{2.71828} = .36788$ and therefore, when $t = \frac{L}{R}$

$$i = \frac{E}{R} (1 - .36788) \quad \dots \quad 7a \text{ (current growing)}$$

$$i = \frac{E}{R} \times .36788 \quad \dots \quad 7b \text{ (current decaying)}$$

showing that when growing, the current will reach the value $.63212 \frac{E}{R}$ in the time $\frac{L}{R}$, while when decaying the current will fall to $.36788$ of its initial value in the time $\frac{L}{R}$.

CHAPTER III.—MEASURING INSTRUMENTS

MEASUREMENT OF DIRECT CURRENTS AND CONSTANT VOLTAGES

1. The electrical quantities which commonly require measurement in D.C. circuits are (i) the current flowing, (ii) the P.D. between various points in a circuit, (iii) the rate at which energy is being converted, i.e. power being supplied or consumed, (iv) the total energy supplied over a given period. Instruments used for measuring current are called ammeters, special forms used for the measurement of small currents being termed milliammeters and microammeters. Instruments used for the measurement of P.D. are called voltmeters, millivoltmeters being in occasional use. The measurement of power is accomplished by the employment of the watt-meter, while instruments which measure the total energy supplied or consumed over a given period are termed energy meters or watt-hour meters. In this chapter it is proposed to describe certain types of all these instruments, chiefly from the aspect of D.C. measurement, although in certain instances the instrument is suitable for use in either A.C. or D.C. circuits.

Requirements of ammeters and voltmeters

2. The following properties are desirable in these instruments, whether used for D.C. or A.C.

(i) They should be "dead-beat" in action. The term dead-beat means that when a given current passes through the meter, or a given P.D. is applied to it, the pointer should immediately register the correct reading and not oscillate on either side of it for an appreciable period. A dead-beat instrument is said to be efficiently "damped", and the steps taken to ensure this in different types of instruments are detailed later.

(ii) They should retain the accuracy of their calibration, always returning to zero when no current flows, and having inappreciable error if used at a temperature differing from that at which calibration was performed.

(iii) The scale should be divided uniformly.

(iv) External electric or magnetic fields should not influence the deflection.

The British Standards Institution recognises three grades of ammeters:—Sub-standard instruments, which are used for calibration purposes, and have an accuracy within ± 0.5 per cent. 1st Grade, and 2nd Grade, with permissible errors, depending upon the working conditions, of from 1 to 2 per cent. and 2 to 4 per cent. respectively. Voltmeters are similarly graded, but the tolerated errors are smaller since the current flowing, and therefore the heat developed, is less. Grade I and Grade II instruments may be calibrated by comparison with sub-standards, and are used for ordinary switchboard requirements.

3. (i) We have already seen that the effects of a current are chemical, thermal and magnetic. Ammeters and voltmeters may be made to operate by any one of these effects. The chemical effect is however ill-adapted for use in a direct reading or "deflectional" instrument, i.e. one which is provided with pointer and scale. Practical direct reading ammeters therefore make use of either the thermal or the magnetic effect.

(ii) The simplest form of current-indicating instrument is the galvanometer which has already been described. The principle used therein, namely the deflection of a suspended or pivoted bar magnet by the field produced by an electric current, is rarely if ever adopted when actual current measurement is required, but the moving coil instrument, presently to be described, is really an inversion of the principle, a powerful fixed magnet causing the deflection of a current-carrying coil. Before proceeding further it is desirable to emphasise that instruments depending upon the deflection of a magnet by a current-carrying coil or vice versa, measure the average value of the current, and give no deflection when an alternating current flows through the conductor. Taking the simple galvanometer as an example, a moment's reflection will shew that if the current is alternating, the magnetic needle will tend to oscillate, turning alternately

CHAPTER III.—PARAS. 4-5

to the left or right hand as the current changes its direction. The inertia of the needle is too great to allow it to perform such changes, at any rate if the change in the direction of the current occurs more than once or twice per second, and therefore the needle remains stationary when such a current is applied to the coil.

(iii) Instruments which depend upon the thermal effect of a current, however, actually measure the rate at which heat is being produced in the instrument, and the deflection of the pointer is proportional to the power dissipated in the instrument, and therefore to the square of the current. From fundamental principles it is apparent that the heat developed in any conductor is independent of the direction of the current, and depends only upon its magnitude. The "square law" above mentioned is the mathematical expression of this physical fact, for the square of the current is always positive, irrespective of the sign, positive or negative, which we arbitrarily assign to its direction. In general, it may be stated that instruments which give a deflection proportional to the average value of the current are suitable for use in direct current circuits only, while those which give a deflection proportional to the mean value of the square of the current are suitable for measurement of either direct, alternating or pulsating currents.

Fundamental principles of instrument design

4. Certain fundamental considerations apply to all types of indicating instruments. In every such instrument there is a moving part of light and therefore delicate construction which is pivoted in jewelled bearings, and this moving member carries the pointer, which moves over a graduated scale. The pointer must be extremely light and yet very rigid, and is generally made from thin aluminium sheet which is drawn into a channel section to attain the desired rigidity, while another design utilises thin aluminium wire which is built up into a rigid structure by cross bracing, hence it is usually called the "girder" type of pointer. If instruments must be opened for repair, the utmost care must be exercised to avoid damage to any part of the moving member. The current passing through the meter, which is proportional to the current or voltage to be measured, sets up a torque or turning moment which causes the moving part to rotate. This actuating torque always increases with the current passing through the instrument, but is not necessarily in strict proportion thereto, although it is so in the moving coil instrument if the magnetic field is correctly distributed. In order to obtain a steady deflection for a given current through the instrument, some opposing torque must be introduced by the rotation of the moving part, and this, which is called the restoring torque because it tends to return the pointer to its initial position, must increase with the angle through which deflection occurs. The pointer will then come to rest at some definite position, called the position of equilibrium, at which the actuating torque and restoring torque are equal.

5. (i) The restoring torque is normally provided by either of two methods, viz : gravity control or spring control. Gravity control is achieved by attaching to the moving member a light arm carrying a small weight ; the arrangement is such that the weight occupies its lowest possible position when the pointer is at the zero of its scale. When the pointer is deflected the arm carrying the weight is inclined to the vertical, and the force of gravity acts upon the weight, producing a torque which tends to turn the pointer to zero. The restoring torque is shown by the diagram (fig. 1), to be equal to wx units, the unit of force usually used being the gram, and the unit of length the centimetre. Hence the restoring force may be said to be wx gram-centimetres. From the diagram it will be seen that if the weight is mounted r centimetres from the pivot, $x = r \sin \theta$ where θ is the angle through which it has been deflected. The restoring force is equal to $wr \sin \theta$ gram-centimetres, and is therefore proportional to $\sin \theta$, and the scale division varies accordingly. The effect of this is to limit the useful range of movement of the pointer to about 80° . Gravity control is most commonly applied in moving iron instruments of the switchboard type.

(ii) The second method, spring control, is invariably used in the design of moving coil instruments. The springs used are flat spiral springs similar to the hair spring which controls the balance wheel of a watch, and the torque exerted by this form of spring is directly proportional to the angle through which it is twisted, hence this spring control gives a restoring torque which is

directly proportional to the angle through which the pointer is deflected. In certain A.C. meters in which this form of control is used, the pointer has a movement of 300° , but in ordinary designs about 120° is rarely exceeded.

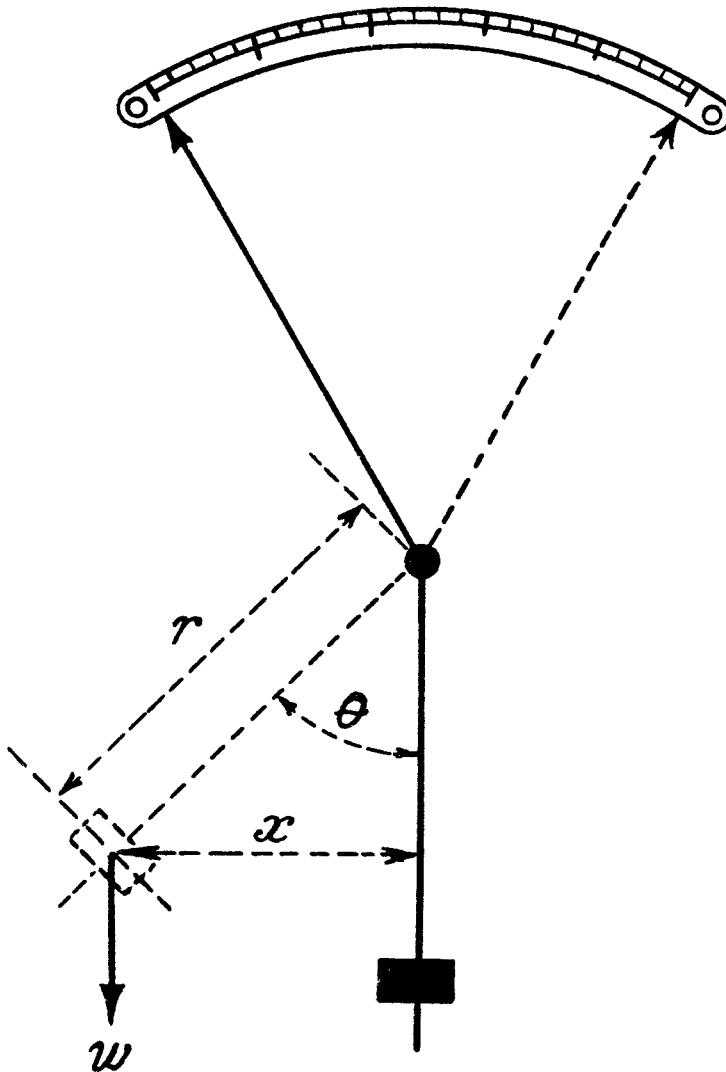


FIG. 1, CHAP. III.—Gravity control.

Moving coil instrument

6. A coil carrying a current is equivalent to a magnet and if placed in a magnetic field will tend to turn into the position in which it embraces maximum flux. This principle is used in the construction of moving coil instruments, as shewn diagrammatically in fig. 2. The magnetic field is provided by a permanent magnet (A), and a soft iron core (B) is mounted between its poles, this serving to concentrate the flux in the air gap, and to cause it to pass through the gap radially. The sides of the moving coil (C) are then always normal to the flux, no matter in what

CHAPTER III.—PARA. 7

position the coil may be, and the deflection is truly proportional to the current in the coil. The coil is wound on a light metal former, and current is led to and from the windings by spiral springs (S) which also act as the controlling device. The restoring force of these springs is proportional to the angle through which the coil has turned from its zero position. When current flows, therefore, the coil takes up a position of equilibrium, in which the torque due to the current is exactly balanced by the resisting torque of the spring. The instrument thus has an evenly divided scale and fulfils requirement (iii).

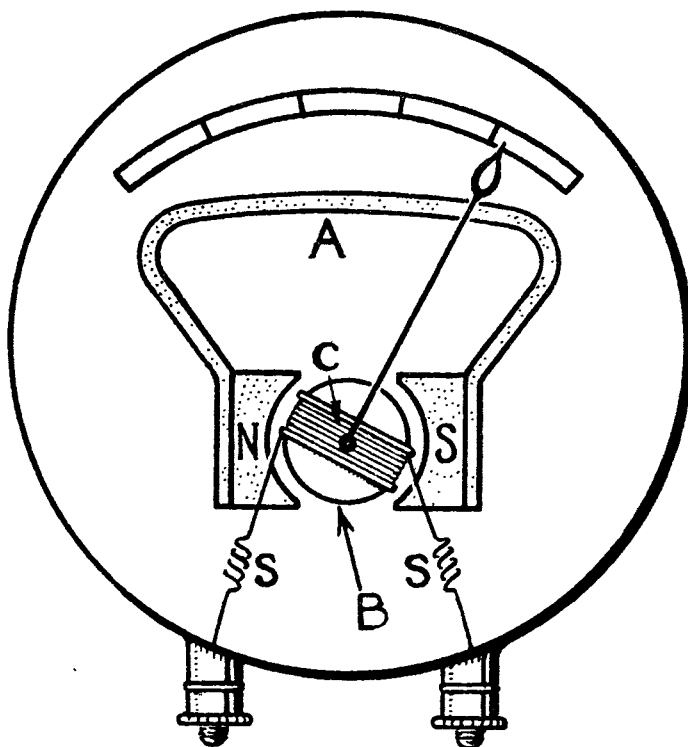


FIG. 2, CHAP. III.—Principle of moving coil instrument.

When in motion the metal former has induced in it an E.M.F. which by Lenz's law tends to oppose the motion which produced it, and thus acts as a damping device. The lightness of coil and pointer also contribute towards the achievement of effective damping. Requirement (iv) is also nearly fulfilled, for as the magnetic field set up by the permanent magnet is very strong, external magnetic fields have little effect. Nevertheless, such an instrument should not be mounted too near to a dynamo or motor. The zero of the instrument is practically unaffected by temperature, and since the permanent magnet and the controlling springs retain their original properties over a long period, the calibration remains very nearly constant. Moving coil instruments can therefore be obtained as sub-standard or any lower grade.

Moving iron instrument

7. (i) Though not as accurate as the moving coil instrument, this type is of simpler and more robust construction. Two kinds are in general use, the attraction type and the repulsion type. Fig. 3 shews the principle of the attraction type. (A) is a fixed coil carrying the current to be measured, and (B) is a disc of soft iron eccentrically pivoted and carrying a pointer. When a

current flows in (A), (B) becomes magnetised and being pivoted eccentrically moves into the coil, causing a deflection of the pointer. The controlling torque is usually provided by a spring (not shewn). Air damping is also provided by a piston (C) moving in a cylinder (D). The deflection is proportional to the square of the current, and the scale is therefore cramped at its lower end.

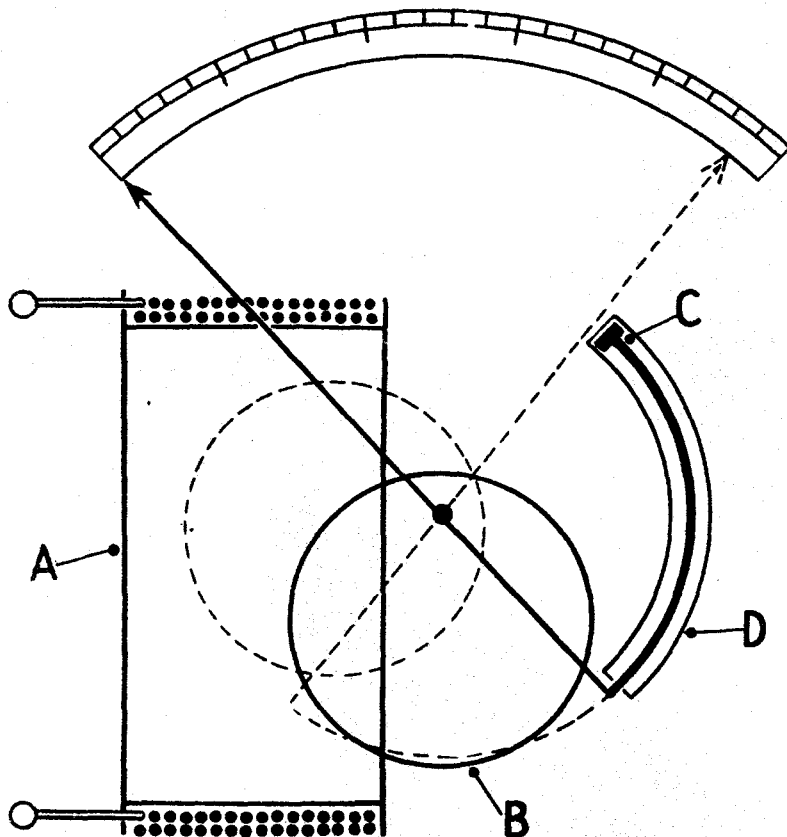


FIG. 3, CHAP. III.—Principle of moving iron instrument (attraction type).

(ii) The principle of the repulsion type is exhibited by fig. 4. Two pieces of soft iron are arranged inside and parallel to the axis of a circular magnetising coil (A) carrying the current to be measured. One of the pieces (B) is of uniform breadth and is attached to a pivoted spindle carrying the pointer. The other piece (C), which is fixed to the case, is curved to a circular arc and is tapered in breadth. Under the magnetising force of the current both pieces of iron become similarly magnetised, and the smaller piece will be repelled from the wide to the narrow end of the larger, the movement being controlled by a flat hair-spring (D). The deflection will be proportional to the square of the current, and consequently the scale will not be uniformly divided. Moving iron instruments are subject to influence by external fields, which however can be much reduced by enclosure within an iron case. Their readings are also subject to error due to hysteresis.

Hot-wire instrument

8. This is shewn diagrammatically in fig. 5. (A) is a wire of high melting point and high specific resistance, platinum-silver or iridium-platinum being commonly employed. One end of

CHAPTER III.—PARA. 9

a phosphor-bronze wire (B) is attached to the centre of the wire (A), the other end being attached to an insulated post (C). A silk fibre is attached to (B), passing round a small pulley (D) which carries a pointer, and is kept in tension by a spring (E). One end of the wire (A) is connected to the metal base plate at (F), the other to an insulated post (G). The action of the instrument is as follows:—When a current flows through (A), the wire expands and sags, this sag being taken up by the wire (B) and the spring (E), thus rotating the pulley (D) so that the pointer moves over the graduated scale. The heating effect of the current is proportional to the square of the current, so that the expansion and consequently the pointer deflection, also varies as I^2 . The scale is therefore cramped at the lower end and requirement (iii) is not fulfilled by hot-wire instruments. They also tend to have an uncertain zero error, due to the wire failing to return to room temperature for a considerable period after current has ceased to flow, while the length

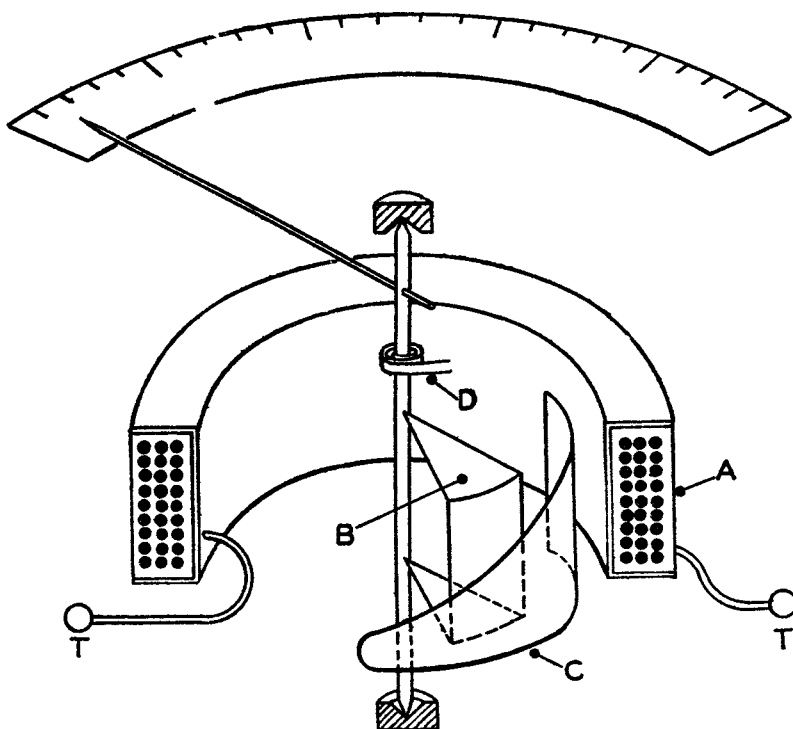


FIG. 4, CHAP. III.—Principle of moving iron instrument (repulsion type).

of time taken to heat the wire to a steady temperature renders the instrument sluggish in action. A simple device which alters the tension of the wire (A) in order to correct the zero error is shewn at (F) in fig. 5. Hot-wire instruments are regarded only as 2nd Grade instruments, and are rarely used for D.C. measurements. Damping is introduced into these instruments by a metal disc (H), carried on the spindle of the pulley (D). When the pulley rotates the disc moves between the poles of a permanent magnet (J). Eddy currents are thus induced in the disc, and these eddy currents inter-acting with the magnetic field, tend to turn the disc in the opposite direction to that in which it is moving.

Connections for use as ammeter or voltmeter

9. The foregoing pieces of apparatus were referred to as “instruments”, because the principles embodied therein can be applied to the measurement of either P.D. or current. The

difference between an ammeter and a voltmeter of the same type is entirely due to their different functions. An ammeter is required to measure current, and must be inserted in series with the circuit in which the measurement is to be made. It must therefore be of low resistance compared with the remainder of the circuit, otherwise the insertion of the ammeter will alter appreciably the current flowing, and also lead to a loss of power in the instrument itself. A voltmeter measures P.D., and must be connected between the points whose P.D. is required. It must therefore have a high resistance so that the extra current taken by the instrument will not lower the P.D. appreciably, and also to reduce the power loss in the instrument itself. In practice the same instrument may be used either as a voltmeter or an ammeter, in combination with suitable resistances. It must of course be calibrated in accordance with the use for which it is intended. Taking the moving coil type as an example, it is usual to manufacture a standard instrument which

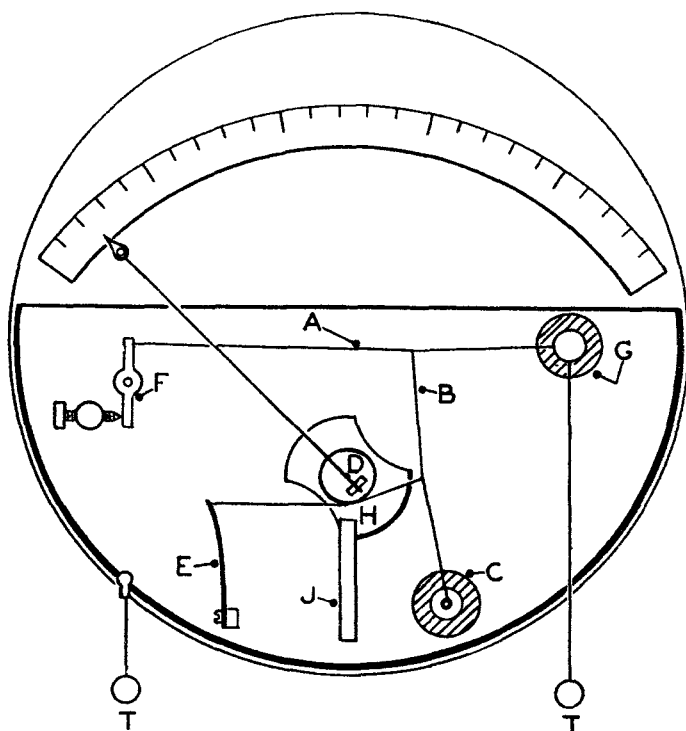


FIG. 5, CHAP. III.—Principle of hot wire instrument.

usually requires about $\cdot 01$ ampere to give full scale deflection, the total resistance of the instrument being say 5 ohms. If it is required to measure up to 10 amperes, it is necessary to connect a resistance in parallel with the instrument of such a value that 9.99 amperes will flow in the parallel resistance and only $\cdot 01$ ampere through the moving coil. A resistance used in this way is called an ammeter shunt (fig. 6a). The correct value of the shunt resistance is easily found, for the P.D. between its ends, and the current flowing in it are both known. As the instrument has a resistance of 5 ohms and takes a current of $\cdot 01$ ampere, the P.D. must be $\cdot 05$ volt, and the resistance of the shunt is $\frac{\cdot 05}{9.9} = \cdot 00505$ ohms. An ammeter shunt usually consists of one or more rectangular sheets of manganin, the ends being hard-soldered into heavy copper blocks to which the instrument leads are attached. For currents up to about 20 amperes the shunt is generally contained in the case of the instrument and this practice is of course compulsory in portable

CHAPTER III.—PARAS. 10-11

instruments. For switchboard use the shunt may be fitted externally, the connecting leads being also supplied. It is essential that these should not be shortened or the calibration may be seriously affected.

10. When an instrument is to be used as a voltmeter, a series resistance must be used with it (fig. 6b). Suppose it is required to use the above instrument to measure P.D. up to 240 volts. The total resistance must then be such that 240 volts will just cause a current of $\cdot 01$ ampere, and the total resistance of the voltmeter must be 24,000 ohms. As the instrument has a resistance of only 5 ohms, it is necessary to connect a resistance of 23,995 ohms in series with it. The appropriate resistance is incorporated in the case of portable instruments but is often mounted separately if the voltmeter is designed for switchboard use. It is usually of eureka wire wound on flat mica strips, or sometimes upon a porcelain or hard-wood bobbin.

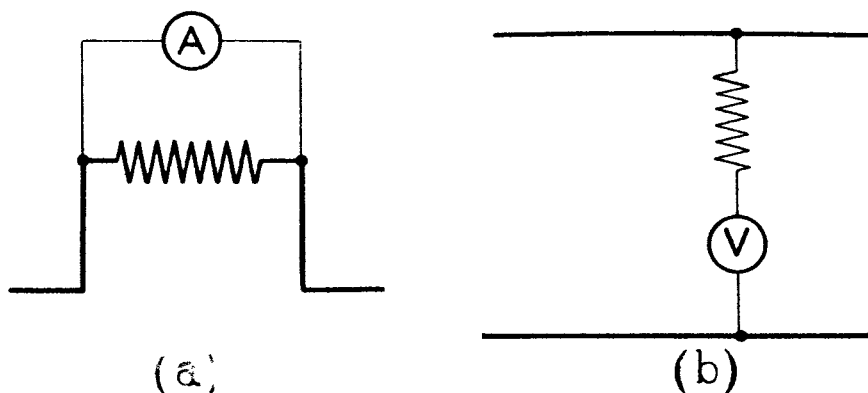


FIG. 6, CHAP. III.—Connection of instrument as ammeter or voltmeter.

A point of interest arises with regard to the power consumed by a voltmeter when used for the measurements of very high voltages, 1,000 volts and upwards. The moving coil instrument previously considered if used on a supply voltage of 1,000 volts, will consume $1,000 \text{ volts} \times \cdot 01 \text{ ampere}$ or 10 watts. In many instances this may be of the same order as the power supplied to the actual load, for example, an aircraft W/T transmitter may take only 90 watts, and the power consumed by the voltmeter is 10 per cent. of the total power supplied. If it is proposed to use a hot-wire instrument conditions are even worse, because it is difficult to design this form of instrument to give full-scale deflection with less than about $\cdot 15$ ampere; at 1,000 volts the voltmeter would then consume 150 watts which is nearly double that required by the transmitter. For this reason the hot-wire voltmeter is rarely used, the development of the electrostatic voltmeter from a laboratory instrument into a form suitable for service and commercial use having provided a much more efficient instrument.

The electrostatic voltmeter

11. The principle of electrostatic attraction cannot be applied to the measurement of current, but is frequently used for measurement of P.D. A typical instrument is shown in fig. 7. A light metal vane (A) is mounted on a pivoted spindle which also carries the pointer. This vane is free to move between two metal plates (B), which are electrically connected; for clarity only one of these plates is shown in the diagram. The moving vane is in metallic connection with the base plate via its controlling spring (not shown), while the plates (B) are insulated from it. When a P.D. is established between the moving vane and the fixed plates, the two systems acquire equal and opposite charges, and the resulting electrostatic attraction causes the vane to move into the space between the plates, and consequently a deflection of the pointer. This deflection is proportional to the square of the P.D. and the graduation of the scale is not uniform. The electrostatic voltmeter carries no current and consequently it consumes no power. It is more easily

and cheaply designed for high voltages—500 volts and above—than for low, and it is consequently used for the measurement of anode voltages in radio transmitters almost to the exclusion of other types.

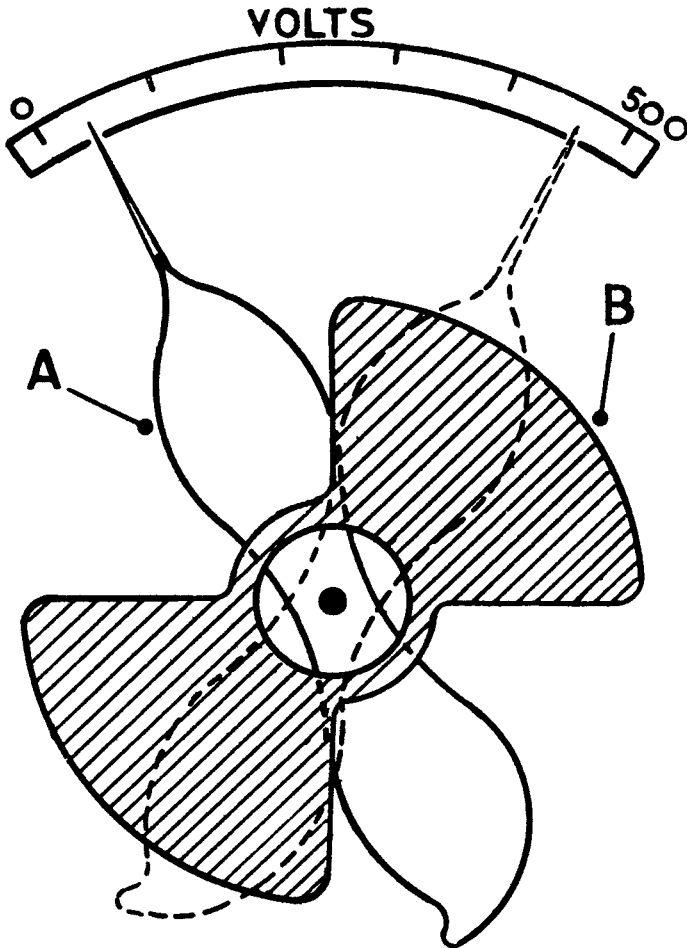


FIG. 7, CHAP. III.—Principle of electrostatic voltmeter.

The thermo-ammeter

12. This name is given in the service to an instrument which depends for its action upon the small E.M.F. which is generated when a junction of two dissimilar metals is heated, and is shewn in fig. 8. The junction is made by spot welding two very fine wires of dissimilar conducting materials at the point at which they make contact with a third wire which serves as a heating device. The thermo-electric couple, as it is called, frequently consists of copper and eureka or copper and constantan wires of about 50 s.w.g. The E.M.F. generated in the thermocouple causes a small current to flow, this current being measured by a low-reading moving coil milliammeter, and thus the complete thermo-ammeter consists of a moving coil milliammeter and thermocouple, the two parts being sometimes mounted separately, while in other designs the thermocouple forms an integral part of the instrument. These instruments are expensive in first cost and their repair entails highly-skilled workmanship. If it is necessary to make *ex tempore*

CHAPTER III.—PARA. 13

current measurements with an instrument of this type, great care must be exercised to ensure that the current-carrying capacity of the thermo-junction is not exceeded, for owing to the fineness of the wires they are incapable of withstanding even a small over-load. This principle is rarely if ever applied in measurement of voltage.

Measurement of resistance

13. The measurement of an unknown resistance by means of the Wheatstone's bridge has already been explained in Chapter I. It is often desirable, however, to make use of an instrument

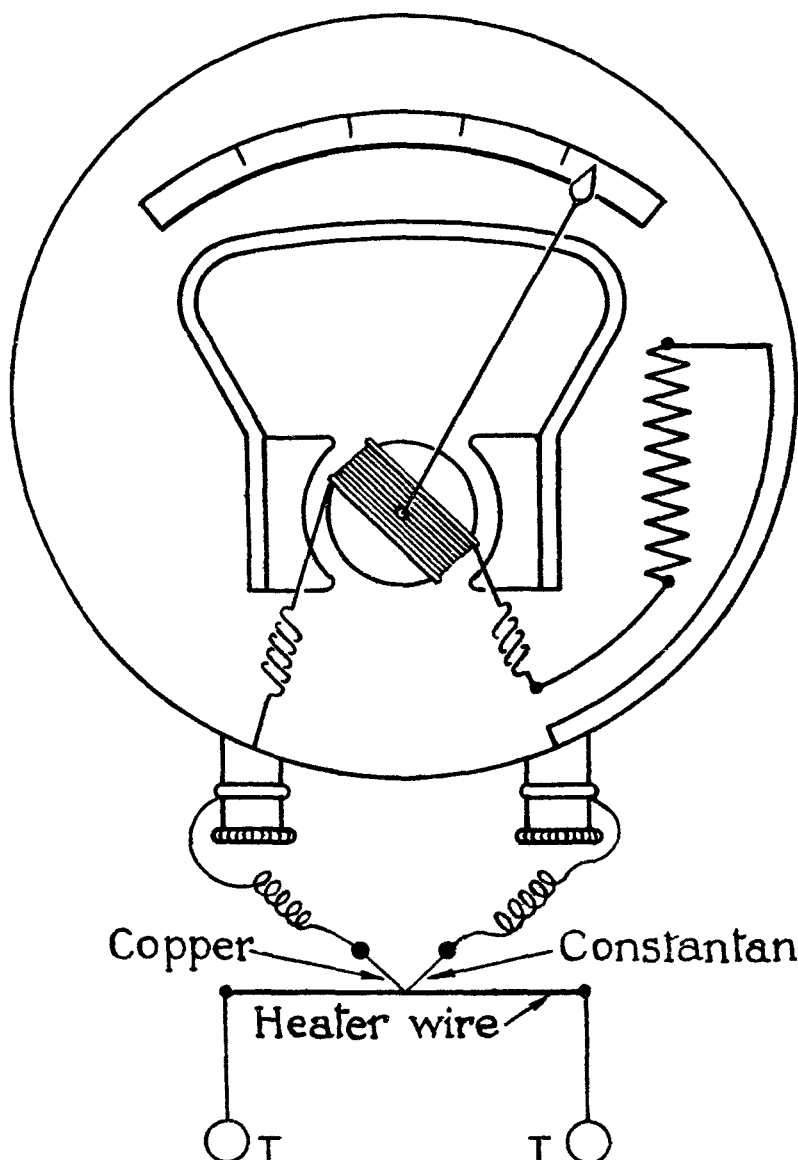


FIG. 8, CHAP. III.—Principle of thermo-ammeter.

which will rapidly determine the value of an unknown resistance, even at the expense of a high order of accuracy; it is preferable that the instrument used for this purpose should be portable, and should be direct reading, i.e. it should be equipped with a pointer and scale, giving the resistance of the circuit under measurement directly in ohms. Two types of instrument are in use for this purpose, the first consisting merely of a moving coil milliammeter in series with a small electric battery, which is usually made up of one or more dry cells and is fitted inside the case of the instrument. Assuming that the E.M.F. of the cell remains constant and its internal resistance negligible, the scale of the milliammeter may be calibrated in ohms instead of in milliamperes, since the resistance and current are in inverse proportion. This type of instrument is not suitable for the measurement of resistances below 1,000 ohms, and high accuracy of measurement of higher resistance cannot be expected. Its great advantage is that by suitable connections and internal arrangement it becomes a universal instrument reading several ranges of volts, amperes and ohms.

14. The true ohmmeter (fig. 9) consists of a special form of moving coil instrument, having a permanent magnet field in which a moving element is free to rotate through an angle of about 100 degrees. This moving element consists of two coils (C), (P), which are mounted at right angles to each other, current being carried to each coil by fine and extremely flexible leads, or by very

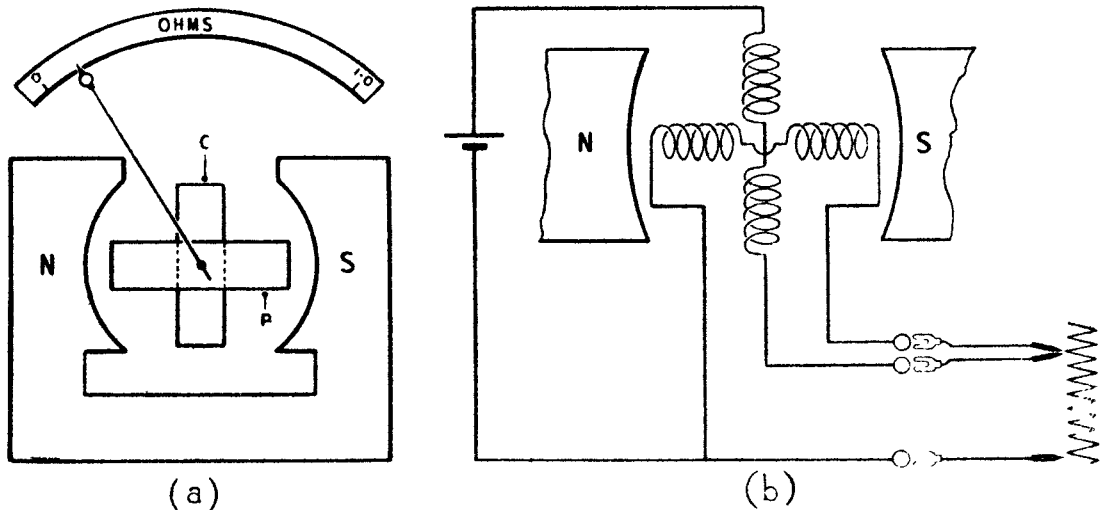


FIG. 9, CHAP. III.—Principle of ohmmeter.

weak springs so arranged that they exert no controlling torque upon the coils. The latter are therefore free, and when not in use the pointer carried by the moving element may rest at any point on the scale. It is important to realise that this is an essential feature of the instrument, and does not signify that it is defective. One of the coils, which is of low resistance and is called the current coil, is connected in series with a source of steady E.M.F. and by suitable leads and terminals to the resistance under test; thus the source of E.M.F., the current coil and resistance under test form a closed conductive circuit. The second coil, which is called the pressure coil, has a resistance which is large compared to the resistance under test, and is connected in parallel with the latter, so that the terminal P.D. of the pressure coil is the same as the voltage drop across the resistance. A typical circuit is given in fig. 9 which shews a type of ohmmeter suitable for the measurement of very low resistances, of the order of $\cdot 05$ ohms.

In this instrument the source of E.M.F. is a single secondary cell and the action is as follows. The current coil carries a current which is inversely proportional to the unknown resistance, while the pressure coil carries a current proportional to the P.D. between the ends of the

CHAPTER III.—PARAS. 15-17

resistance. The direction of the current through the two coils is such that the torque due to each is in opposition, thus a large current through the current coil tends to turn the coil so that the pointer reads zero, while if no current passes through the current coil, i.e. if the resistance under test is infinitely large, only the torque of the pressure coil is operative, and the pointer is deflected vigorously to the high reading end of the scale. For any value of resistance within the range of the scale, there is some position of equilibrium between the two opposing torques and the pointer comes to rest at a definite point on the scale, which can therefore be calibrated directly in ohms. It must be observed that this equilibrium position is not dependent upon the E.M.F. applied, because both the current coil torque and the pressure coil torque are varied to the same degree by a change in voltage.

15. For measurement of very high resistances, such as the insulation between an electric conductor and earth, the ohmmeter is of the same general form as the above, but the E.M.F. is necessarily higher, and may be 250 volts, 500 volts or in extreme cases 2,000 volts. It is then usual to use a source of supply consisting of a hand driven direct current generator which is often contained in the same casing as the ohmmeter proper and may derive its excitation (*see* Chapter IV) from the permanent magnetic field of the ohmmeter. In order to prevent excessive currents from flowing in the current coil in the event of the unknown resistance being abnormally low a protecting coil of the order of 100,000 ohms resistance is usually fitted in series with it. When the circuit under test has considerable capacitance, it is essential that the applied E.M.F. shall not fluctuate, otherwise the P.D. across the resistance will vary with the state of charge or discharge of the capacitance, and therefore it is preferable to arrange the generator drive through a slipping clutch, so that provided the speed of rotation exceeds a certain number of revolutions per minute, the generator speed remains constant. This arrangement is a feature of the "megger" testing set, but is not fitted in the "meg tester". When testing circuits of considerable capacitance with the latter instrument, care must be taken to maintain the generator speed as constant as possible.

POWER MEASUREMENT

16. An instrument which measures the power supplied to or delivered by any piece of apparatus is called a watt-meter. In direct current circuits the same information can be obtained by multiplying the current flowing through the device by the P.D. at its terminals, and as ammeters and voltmeters are generally fitted, this product is easily obtained and the watt-meter becomes redundant. In alternating current circuits, however, power is not as a rule equal merely to the product of volts and amperes, but also depends upon the amount of energy stored in the form of alternating magnetic and electric fields, and is equal to the product of volts and amperes further multiplied by a quantity called the power factor, the maximum value of which is unity and the minimum value zero. The more common forms of watt-meter will be described in Chapter V. Supply meters are used for the purpose of measuring the total energy supplied during a certain period. They are called integrating instruments, which means that they measure the average value of the energy supplied during a very short time and automatically register the sum of all these averages during the interval of operation. Supply meters may be divided into two main classes, viz. ampere-hour meters and watt-hour meters. Ampere-hour meters measure the quantity of electricity which passes, the energy supplied or consumed being derived by assuming that the P.D. remains constant, e.g. V volts. The meter measures the product of current and time, or $I \times t$, and this quantity further multiplied by the P.D. V , is the energy supplied in watt-hours. The dial may therefore be calibrated in watt-hours. Three principal types of ampere-hour meter have been developed, and a representative of each type will be described.

Electrolytic meters

17. A typical specimen is the Bastian meter, which depends for its action on the phenomenon of electrolysis of a very dilute solution of caustic soda. The electrolyte is contained in a glass vessel having a long neck of uniform bore. Two nickel electrodes are fitted in a compartment

made from an insulating material, and insulated leads from these electrodes are carried to the terminals of the instrument, to which the external circuit is connected. The whole of the supply current passes through the electrolyte, which is decomposed into its constituent gases hydrogen and oxygen. These gases escape into the air and the level of the liquid in the neck gradually sinks. The fall of the surface level is proportional to the number of coulombs which pass, and a scale which reads watt-hours, or kilowatt-hours (on the assumption of constant P.D.), is placed parallel to the neck, so that the level of the liquid gives the energy consumed in kilowatt-hours (B.O.T.U.). A thin film of oil on the surface of the liquid prevents evaporation of the electrolyte.

The advantages of the Bastian meter are its accuracy at low loads, owing to the absence of moving parts and consequent friction, its cheapness and simplicity. On the other hand, the gases evolved form an explosive mixture, and it is difficult to read while appreciable current is flowing owing to the formation of gas bubbles on the surface. The fact that a P.D. of about 2 volts occurs in the meter itself also causes it to be wasteful. Thus if a current of 50 amperes is flowing the power wasted in the meter itself is 100 watts or $\cdot 1$ kilowatt, hence 1 B.O.T.U. is consumed by the meter in 10,000 hours.

Commutator motor meters

18. This type of meter contains a member which rotates in the powerful magnetic field of a permanent magnet, fig. 10. This rotating member or armature is wound with wire but contains

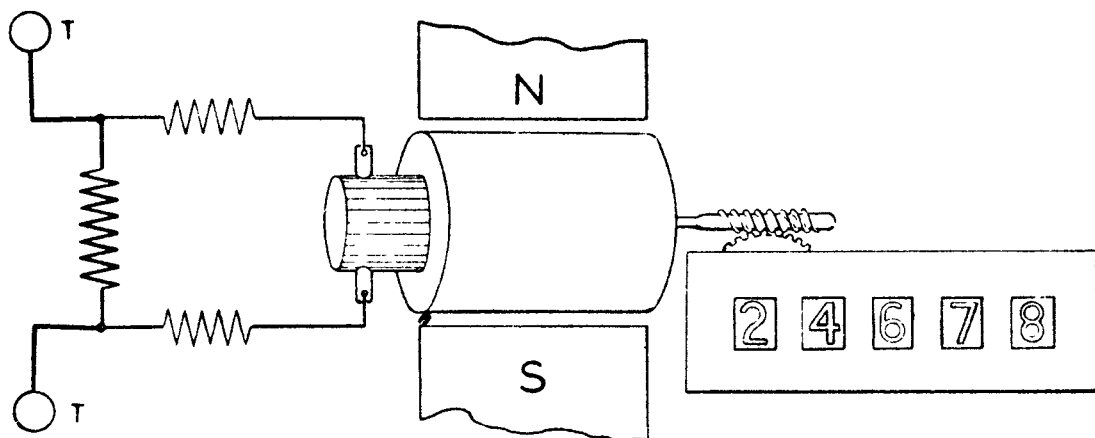


FIG. 10, CHAP. III.—Principle of commutator motor meter.

no iron core, and is very light. It rotates on a hardened steel pivot in jewelled bearings, the steel spindle carrying the armature being extended at one end in order to carry a worm which engages with a worm wheel. The latter is the first of a train of wheels which form a register of the cyclometer type. The rotating armature is connected in series with a certain resistance, and carries only a definite fraction of the whole current, the remainder being carried by the shunt which forms an integral portion of the instrument. The reaction between this current and the permanent magnetic field in which the armature is situated causes rotation of the armature as described in the following chapter. Provided that the friction is negligible, the speed of rotation is directly proportional to the current flowing, and assuming that the applied P.D. is constant, the number of revolutions executed in a given time is directly proportional to the energy supplied. The cyclometer dials are therefore directly calibrated in B.O.T.U.

Mercury motor meters

19. In this type the rotating member consists of a disc of copper, the poles of the permanent magnet which provides the magnetic field being situated in such a position as to embrace only

CHAPTER III.—PARA. 19

a sector of the disc, as illustrated in fig. 11. The disc rotates about a vertical pivot in jewelled bearings, a worm and worm wheel being used to convey the rotation to the counter mechanism as in the previous type described. The chamber (B) surrounding the disc is of insulating material and is filled with mercury, (A). Current is led into the mercury by suitable connections, passing through the copper disc and out by the lower bearing of the spindle, the jewel bearing of which is suitably arranged for this purpose. The direction of current in the disc is approximately radial, and application of the left-hand rule will give the direction in which the disc tends to rotate, while eddy currents induced in the disc owing to its rotation (the direction

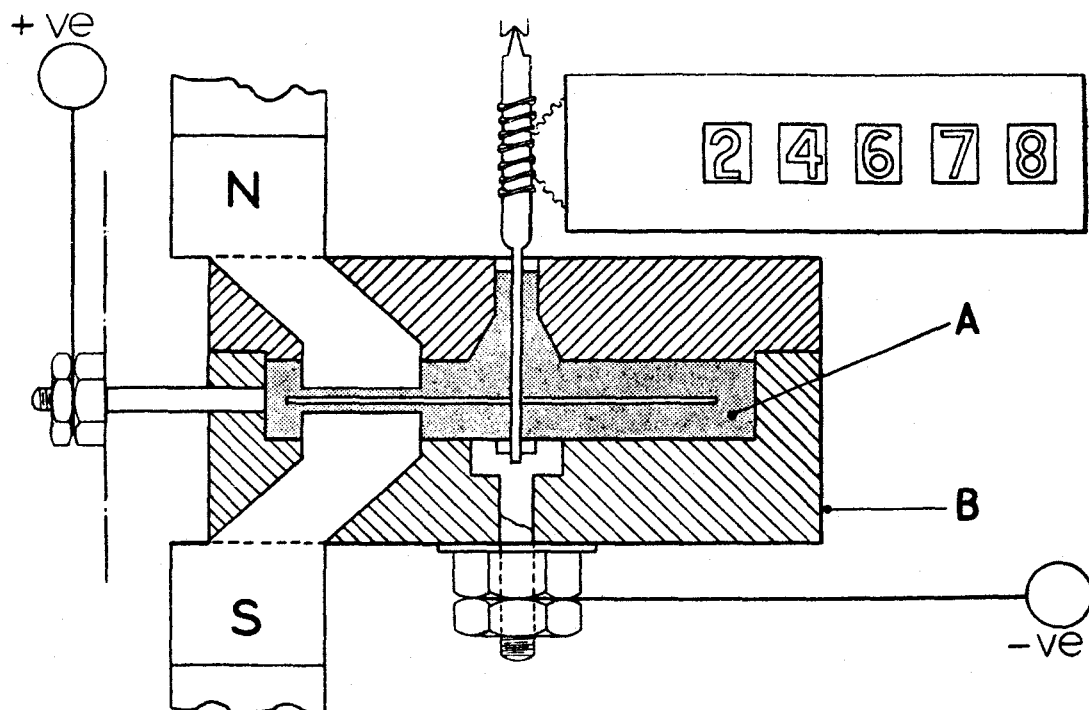


FIG. 11, CHAP. III.—Principle of mercury motor meter.

of which can be found by the right-hand rule) tend to retard the motion. The result of the combined phenomena is that the disc rotates at a speed which is directly proportional to the current flowing, and with constant P.D. this is proportional to the power supplied at any instant. Over a period of time the number of revolutions, as indicated by the cyclometer, is proportional to the number of B.O.T.U. which have passed during the period, and the dials are graduated directly in B.O.T.U.

True energy or watt-hour meters are not often used for D.C. supply, except at central power stations where the total output current is very large, and it is not necessary to describe them here. Such types of watt-hour meter as are likely to be met with in the service are described in Chapter V.

CHAPTER IV.—DYNAMO-ELECTRIC MACHINERY

Introductory

1. In Chapter I it was stated that an electromotive force may be produced by any one of four methods, namely, chemical, thermal, frictional or electro-magnetic. Of these, the only one which lends itself to the conversion of mechanical into electrical energy on a large scale is the latter, and it is now proposed to discuss the means by which this conversion is achieved. We have seen that whenever relative motion takes place between a conductor and a magnetic field, the change of flux-linkage with the conductor results in the production of an E.M.F. The form of relative motion which is most readily adapted to this end is the rotation, either of a conductor in a stationary magnetic field, or of a magnetic field in the neighbourhood of a stationary conductor. The simplest instance is provided by the rotation of a closed conductive loop in a uniform field, which results in the production of an alternating E.M.F. Before proceeding further it is necessary to introduce certain definitions used in connection with alternating quantities.

Definitions

2. An alternating quantity may be briefly defined as one which periodically reverses its direction. Such quantities possess three characteristics by which they may be completely described. They are :—

- (i) Frequency.
- (ii) Wave form.
- (iii) Amplitude or peak value.

In order to define these characteristics let the alternating quantity be an electromotive force acting in an electric circuit. At a given instant the E.M.F. may be supposed to be zero and this instant will be called zero time. After 0·001 second, the E.M.F. may have a value of 50 volts acting in a certain direction in the circuit, after 0·002 second, a value of 95 volts in the same direction, and so on through a series of values first increasing and then decreasing, again reaching zero value after 0·005 seconds. The E.M.F. then commences to grow in magnitude but in the opposite direction, passing through a series of values as before until it reaches again zero value. The complete series of values takes 0·01 seconds to perform, and the variation is then repeated again and again for some indefinite period. An E.M.F. of this nature is plotted on a time scale in fig. 1. The value of an alternating quantity at any given instant is called its *instantaneous*

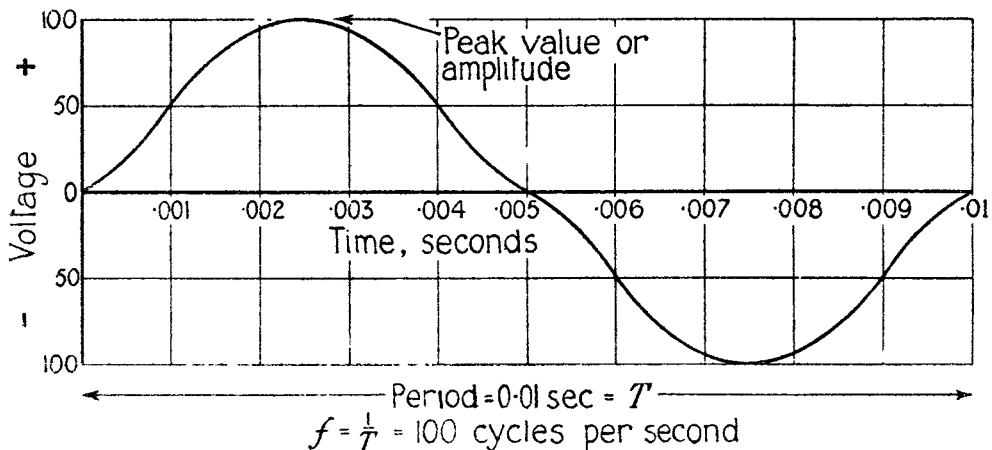


FIG. 1, CHAP. IV.—Graphical representation of an alternating E.M.F.

CHAPTER IV.—PARA. 3

value. A complete series of values, both positive and negative, is called one *cycle*. The time taken for the execution of one cycle is called the *period* or *periodic* time, and is denoted by the symbol T . In the example given above the periodic time is 0.01 second. The *frequency* is the number of cycles completed in one second and is denoted by the symbol f ; since in the above instance, one cycle is performed in 0.01 seconds, there will be $\frac{1}{0.01}$ or 100 cycles per second. The frequency

and the periodic time are thus related by the equation $f = \frac{1}{T}$. The shape of the curve showing the variation of instantaneous value with time is called the *wave form*, and the *amplitude* or *peak value* is the maximum value reached during each half cycle.

Production of an alternating E.M.F.

3. Consider a closed loop P, Q, capable of rotation about an axis X, X', in a uniform field produced by the magnetic poles N, S, fig. 2. This arrangement is shown in cross-section in the following diagram (fig. 3), the two horizontal members of the loop being indicated by circles. The instant when the loop is in the neutral plane between the poles will be taken as zero time, and it will be assumed to rotate in a clockwise direction. At the instant when rotation commences, then, the loop is moving in such a direction that the change of flux linking with it is zero, and no

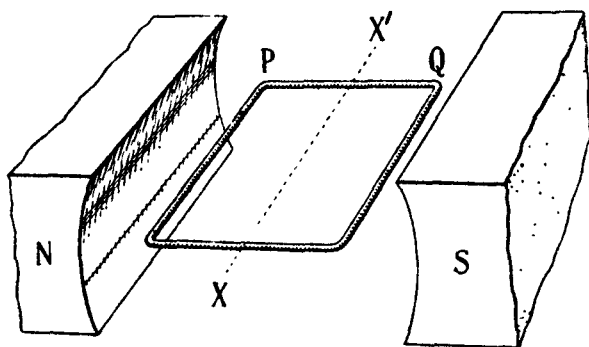


FIG. 2, CHAP. IV.—Closed loop in magnetic field.

E.M.F. is induced, although it will be observed that maximum flux is enclosed at this moment (fig. 3a). After a short interval of time, the loop will reach the position shown in fig. 3b, having turned through an angle of 30° . The conductor P is moving downward, while the conductor Q is moving upward, and the number of tubes of flux linking with the loop is decreasing. An E.M.F. is, therefore, induced in each conductor; its direction may be found by the right hand rule, and is indicated in the conventional manner in the diagram. By Faraday's law the magnitude of the E.M.F. will be proportional to the rate at which the flux-linkage with the loop is changing. In fig. 3c the loop has rotated through 90° from its original position, and no flux whatever is linking with the loop. At first sight it might appear that no E.M.F. will be induced in this position, but this is not so, for actually the flux is in process of complete reversal in direction at this instant, and it will be seen later that the maximum E.M.F. will be generated when the loop is passing through this position. As rotation continues the loop eventually reaches the position indicated in fig. 3d. At this instant the two sides of the loop are moving in a direction parallel to the magnetic flux, and no change of flux-linkage takes place, hence the induced E.M.F. is zero. During the succeeding semi-rotation, the conductor P is moving upward while the conductor Q is moving downward through the magnetic field and the direction of the induced E.M.F. is reversed in each. Maximum induction will again take place when the loop is passing through the position shown in fig. 3e, and finally, after one complete revolution, the loop regains its original position

in which the induced E.M.F. is again zero (fig. 3f). The rotation of the loop through one complete revolution of 360° has thus resulted in the production of a single cycle of alternating E.M.F., and so long as the rotation is continued the foregoing process will be repeated.

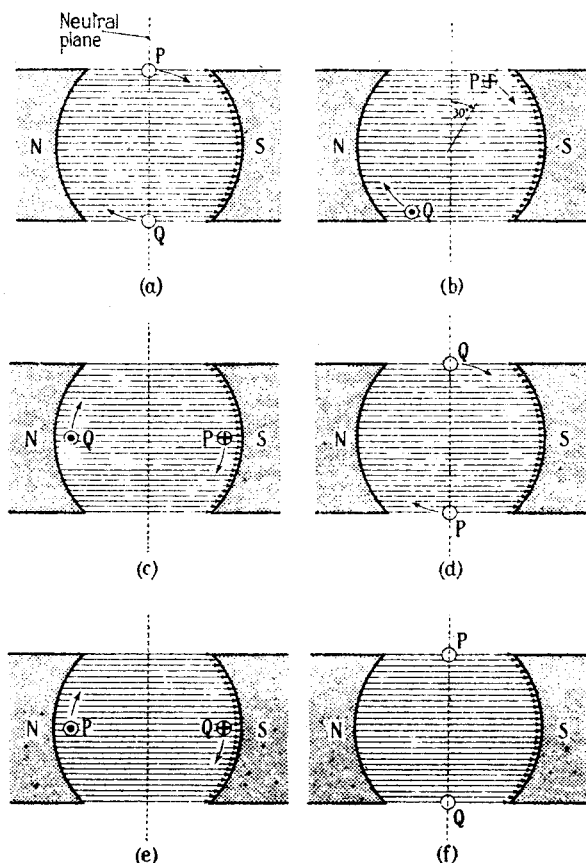


FIG. 3, CHAP. IV.—Production of alternating E.M.F.

Sine law

4. In order to establish an expression giving the E.M.F. induced in the loop throughout its rotation, certain results obtained in Chapter II must be recalled. It was there stated that if a conductor of length l centimetres is moved with a uniform velocity of U centimetres per second through a uniform magnetic field of density B tubes per square centimetre (gauss), the E.M.F. induced is

$$e = BIU \times 10^{-8} \text{ volts.}$$

In the present instance, the conductors are assumed to rotate with constant velocity in space, but the relative motion between conductor and flux depends upon the angle at which the loop is moving through the flux at every instant throughout the revolution. The induced E.M.F. at any instant is proportional to the sine of the angle through which the loop has turned, measured from its initial position in the neutral plane. This may be shown by the aid of fig. 4 in which the magnetic field and the loop PQ have been re-drawn. At the instant depicted, the loop has turned through an angle θ from its initial position, and the direction of motion of the conductor P

CHAPTER IV.—PARA. 5

is shown by the line RP , the length of which may represent the velocity U of the conductor to any convenient scale. The velocity RP may be resolved into two component velocities which are mutually perpendicular, first the component RS which is parallel to the magnetic field; motion in this direction results in no change of flux-linkage and consequently this component of the velocity cannot be responsible for the production of an E.M.F.: second, the component SP which is perpendicular to the flux; motion in this direction results in a change of flux-linkage, and the induced E.M.F. is proportional to the magnitude of this component of the velocity.

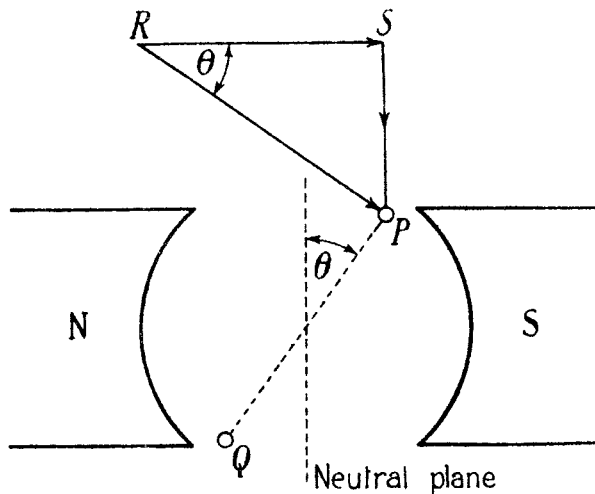


FIG. 4, CHAP. IV.—Components of velocity of P relative to direction of field.

Since RS is perpendicular to the initial plane of the loop and RP is perpendicular to the loop when it has turned through the angle θ , it follows that the angle SRP is also equal to θ , and $SP = RP \sin \theta$; the instantaneous E.M.F. developed in the conductor P is therefore proportional to $RP \sin \theta$, and will reach a maximum when $\theta = 90^\circ$, $\sin \theta = 1$. At this instant the induced E.M.F. is equal to $BIU \times 10^{-8}$ volts, and therefore, if e_r is the E.M.F. induced at any instant in the conductor P ,

$$e_r = BIU \sin \theta \times 10^{-8} \text{ volts.}$$

In a similar manner, it is found that the E.M.F. induced in the conductor Q at any instant is

$$e_q = BIU \sin \theta \times 10^{-8} \text{ volts,}$$

and the total induced E.M.F. is

$$e = e_r + e_q = 2 BIU \sin \theta \times 10^{-8} \text{ volts.}$$

The maximum or peak value $\mathcal{E}$ of the induced E.M.F. is $2 BIU \times 10^{-8}$ volts, and the preceding equation may be written

$$e = \mathcal{E} \sin \theta$$

where

$$\mathcal{E} = 2 BIU \times 10^{-8}$$

Sine curve

5. Having established the manner in which the induced E.M.F. varies from instant to instant during the rotation of the loop, it can be seen from fig. 4 that if we allow RP to represent the peak value $\mathcal{E}$ of the induced E.M.F. to any convenient scale, the line SP represents the instantaneous value. This forms the basis of a convenient method of plotting a curve showing the instantaneous value of the E.M.F. for all values of the angle θ . In fig. 5, the line OP is assumed

to rotate about the point O in a counter-clockwise direction, its length representing the peak value of the E.M.F. to any convenient scale. A horizontal line is drawn through the centre of rotation, O, and upon this a scale of degrees of rotation is set up. As the vertical line PS is equal to $OP \sin \theta$, it represents to scale the instantaneous E.M.F. $e = \mathcal{E} \sin \theta$. An ordinate of length PS is therefore erected on the scale of degrees, having the value of the angle θ as abscissa. The procedure is shown for a particular value of θ , while the dotted lines indicate the manner in which other points are obtained. When the point P lies above the horizontal through the centre of rotation O, the instantaneous E.M.F. is plotted as a positive quantity, i.e. e is regarded as positive for values of between 0° and 180° , while for values between 180° and 360° , e is regarded as of negative sign.

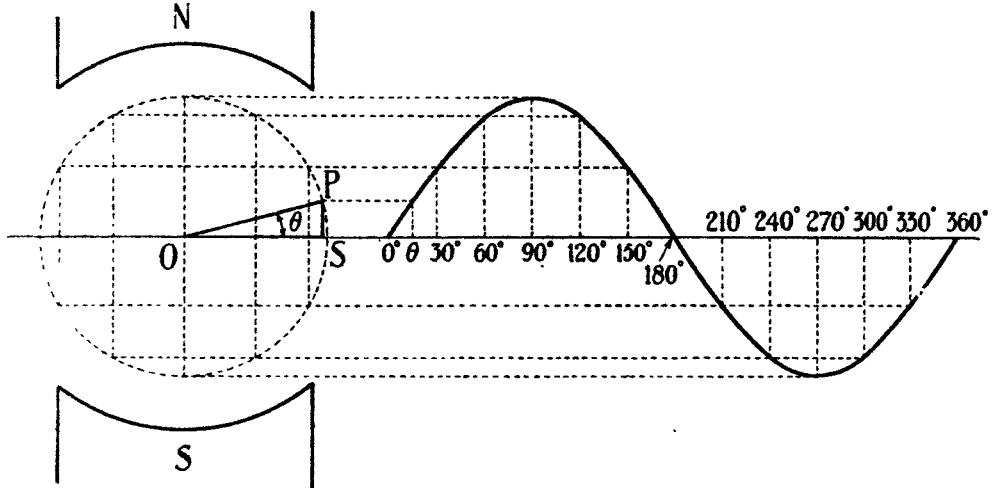


FIG. 5, CHAP. IV.—Sinusoidal E.M.F.

The curve drawn through all the points so obtained is called a sine curve, and an E.M.F. which varies in this way is said to be sinusoidal. This is the simplest wave form which an alternating quantity may assume, and is very nearly approached in a well-designed source of alternating E.M.F. For this reason, it is usually assumed that an alternating E.M.F. is sinusoidal, unless the contrary is distinctly stated.

Angular velocity

6. It is convenient to express the angle θ in such a way that the instantaneous value is made to depend upon time, rather than the angle through which the loop has turned. To do this, it is first necessary to express the angle in radians instead of degrees. A radian is the angle at the centre of a circle subtended by an arc equal in length to the radius of the circle, and is the trigonometrical unit of angular measurement. The circumference of a circle contains 360 degrees, and its length is 2π times its radius, so that a rotation through 360° is the same as rotation through 2π radians.

Hence $360 \text{ degrees} = 2\pi \text{ radians}$

$$1 \text{ degree} = \frac{2\pi}{360} \text{ radians}$$

$$n \text{ degrees} = \frac{n \times \pi}{180} \text{ radians}$$

In performing one revolution, then, the loop passes through an angle of 2π radians. If it rotates at f revolutions per second, it passes through $2\pi f$ radians per second. This is called its angular

CHAPTER IV.—PARA. 7

velocity and is denoted by the greek letter ω (omega). After an interval of t seconds from the commencement of rotation, the loop has rotated through an angle of ωt radians, and the E.M.F. at this instant is

$$e = 2 B l U \sin \omega t \\ = \mathcal{E} \sin \omega t.$$

Example.

A loop of wire two feet square rotates in a uniform magnetic field having a density of 1,000 gauss, at a speed of 600 r.p.m., commencing from a position in the neutral plane. Find the peak value of the induced E.M.F. and the instantaneous value (i) 0.015 seconds and (ii) 1.56 seconds from the commencement of rotation.

Each side of the loop passes round the circumference of a circle 2 feet in diameter, in $\frac{1}{60}$ minute or 0.1 second. The linear velocity of each is therefore

$$U = \frac{\pi d \text{ ft.}}{0.1 \text{ sec.}} = \frac{\pi \times 2 \times 30.48 \text{ cm.}}{0.1 \text{ sec.}} \\ = 1940 \frac{\text{cm.}}{\text{sec.}} \\ \mathcal{E} = 2 B l U \times 10^{-8} \\ = 2 \times 1000 \times 2 \times 30.48 \times 1940 \times 10^{-8} \\ = 2.34 \text{ volts.}$$

The time taken to execute one complete cycle of E.M.F. is 0.1 second and the frequency is $f = \frac{1}{T} = 10$ cycles per second. The angular velocity is therefore $2\pi \times 10$ radians per second.

After rotation for 0.015 second the E.M.F. will be

$$e = 2.34 \sin (62.8 \times 0.015) \\ = 2.34 \sin 0.3\pi \\ = 2.34 \sin 0.942.$$

7. It is now necessary to find the numerical value of $\sin 0.942$. To do this, convert 0.942 radians to circular measure by the relation already given, i.e.

$$0.942 \text{ radians} = 0.3\pi \text{ radians} = 0.3\pi \times \frac{180}{\pi} \text{ degrees};$$

$$\text{or } 0.942 \text{ radians} = 54 \text{ degrees.}$$

From trigonometrical tables, $\sin 54^\circ = 0.8090$

$$\therefore e = 2.34 \times 0.8090 \\ = 1.9 \text{ volts, approximately.}$$

After 1.56 seconds, the loop has rotated through an angle of ωt radians, where

$$\omega t = 2\pi \times 10 \times 1.56 \text{ or } 31.2\pi \text{ radians.}$$

$$31.2\pi \text{ radians} = \frac{31.2\pi \times 180}{\pi} \text{ degrees} \\ = 5616 \text{ degrees.}$$

The loop has therefore executed $\frac{5616}{360}$ or 15.6 revolutions. Since after 15 revolutions the loop is in the neutral plane, we are only concerned with the E.M.F. developed after the ensuing 0.6 of a revolution, i.e. after $0.6 \times 360 = 216$ degrees of rotation from the time at which the loop last

passed through the neutral plane. Trigonometrical tables give the values of $\sin \theta$ for angles between 0° and 90° only, but fig. 5 shows that, for angles between 90° and 180° , $\sin \theta = \sin (180^\circ - \theta)$. Similarly, for angles between 180° and 270° , $\sin \theta = -\sin (\theta - 180^\circ)$ and for angles between 270° and 360° , $\sin \theta = -\sin (360^\circ - \theta)$. For example, the instantaneous E.M.F. after rotation through 210° is of the same magnitude as after rotation through 30° , but is acting in the opposite direction round the loop. Mathematically, this is expressed by writing

$$\sin 210^\circ = -\sin 30^\circ$$

After 1.56 seconds, then, the instantaneous E.M.F., e , is $2.34 \sin 216^\circ$ and $\sin 216^\circ = -\sin 36^\circ$. As $\sin 36^\circ = .5878$,

$$\begin{aligned} e &= -2.34 \times .5878 \\ &= -1.375 \text{ volts.} \end{aligned}$$

8. As the loop forms a complete conductive circuit, a conduction current will be established in it by the alternating E.M.F. The current will also be an alternating quantity, having the same frequency as the E.M.F. By means of devices to be described later, the current is led to an external circuit, where it can perform useful work. The magnitude of the current depends on the opposition offered by the conductor, which is never less, but may be considerably greater than the resistance of the circuit, and is called its impedance. The factors governing the magnitude of the impedance, and the resulting relations between current and voltage in an A.C. circuit are dealt with in Chapter V. The simple arrangement of relative motion between a magnetic field and a conductor described above is prototype of all kinds of dynamo-electric machinery, but many modifications are necessary in order to produce a machine of practical utility some of which are of general application to all types and others peculiar to the particular function which the machine is designed to perform.

Frequency

9. In the arrangement shown in figs. 2 and 3, a complete cycle of E.M.F. is generated by every revolution of the loop. The frequency is therefore equal to the number of revolutions per second. Suppose now that the magnetic field is maintained by two pairs of poles, arranged alternately as shown in fig. 6. One complete cycle of E.M.F. will then be generated in a conductor

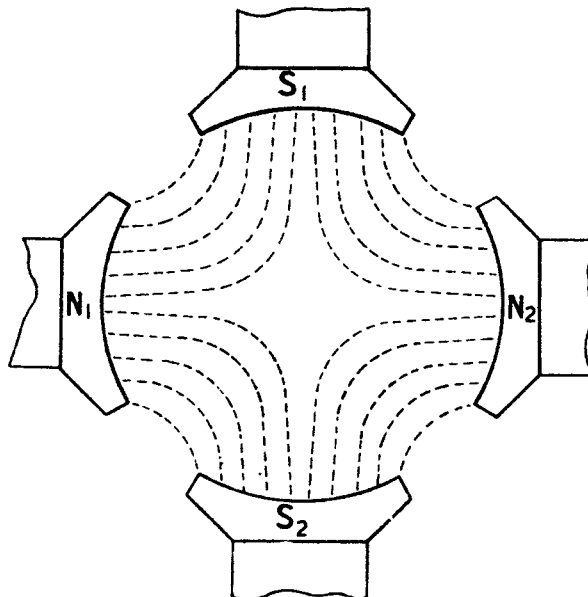


FIG. 6, CHAP. IV.—Four-pole field.

CHAPTER IV.—PARA. 10

passing from N_1 to N_2 , and another cycle in passing from N_2 to N_1 , so that with two pairs of poles two complete cycles are executed in a single rotation of the loop. In general if there are p pairs of poles there will be p cycles per revolution, and the frequency is given by the equation

$$f = p \times \text{number of revolutions per second.}$$

For example, if a twelve-pole machine is required to give 180 cycles per second, it must rotate at $\frac{180}{6} = 30$ revolutions per second or 1,800 r.p.m. With one pair of poles, then, one complete cycle is obtained in 360 degrees of rotation. With a four-pole field one cycle is obtained in 180 degrees of rotation, but the resulting E.M.F. passes through the same variation as with the two-pole arrangement, except that two complete cycles are executed in one revolution. This leads to the conception of "electrical" as opposed to "geometrical" degrees. Each conductor is said to pass through 360 electrical degrees in the space through which a complete cycle of E.M.F. is generated. One complete revolution, in a four-pole field, corresponds to 720° , in an eight-pole field $1,440^\circ$ and so on.

The field magnets and armature

10. Permanent magnets are unsuitable for the production of the powerful field which is generally desirable in alternating current generators, being expensive to manufacture in large sizes, easily damaged by rough treatment, and liable to become demagnetized by the action of the current in the armature winding. The field is, therefore, generally produced by an electro-magnetic system consisting of a cast steel yoke of cylindrical form which forms the carcass of the whole machine, and upon which are radially mounted soft iron pole-pieces which project inwards as shown in fig. 7. Each pole-piece carries a magnetizing winding supplied with current

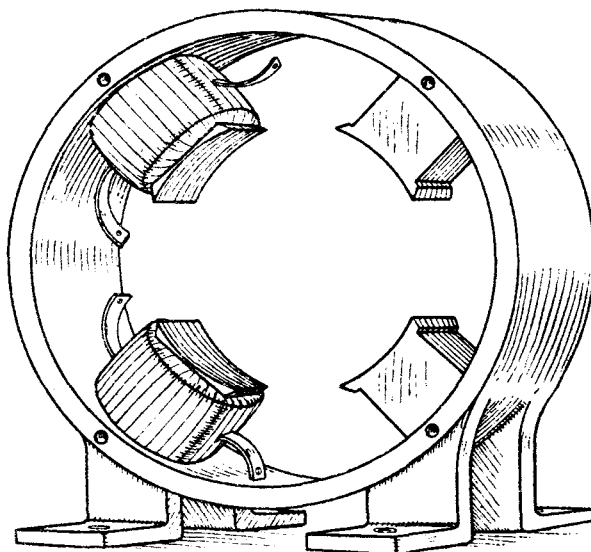


FIG. 7, CHAP. IV.—Carcass with pole pieces and field winding.

from some external source ; two of these coils are shown in position. A further advantage of an electro-magnetic field is the ease with which the flux density can be varied. It will be remembered that the amplitude of the E.M.F. depends directly upon the flux density and speed of rotation. As variation of the speed of rotation varies both the magnitude of the E.M.F. and the frequency, and the latter is generally required to be constant, it is necessary to control the magnitude of the E.M.F. not by speed control but by variation of the flux density. This is easily arranged by means of a resistance in series with the field magnet winding, which controls the magnetizing current and therefore the magnetic field strength ; this control is termed a field regulator.

11. In order to produce an appreciable electromotive force in the rotating loop it is desirable that the flux density of the field shall be as high as possible. This can be achieved by reducing the reluctance of the magnetic circuit, and to do this the air gap in the magnetic field must be reduced to a minimum compatible with free rotation of the loop. In practice the conductors in which the E.M.F. is to be induced may be wound on an iron core of cylindrical form, and the field magnets so shaped as to embrace this core as closely as possible. The conductors are then either wound in longitudinal slots cut in the periphery of the core parallel to the shaft upon which it is mounted, or are pushed through tunnels drilled through the core near and parallel to the surface. The whole assembly including the shaft, core and winding, is termed the armature, and the two types of winding are referred to as slot-winding and tunnel-winding respectively.

Eddy currents

12. The introduction of the iron core gives rise to a new problem. Its rotation in the magnetic field produces varying electromotive forces in the core itself, and since it is of conducting material, currents are set up which circulate in ever-changing paths in the iron, as in fig. 8a.

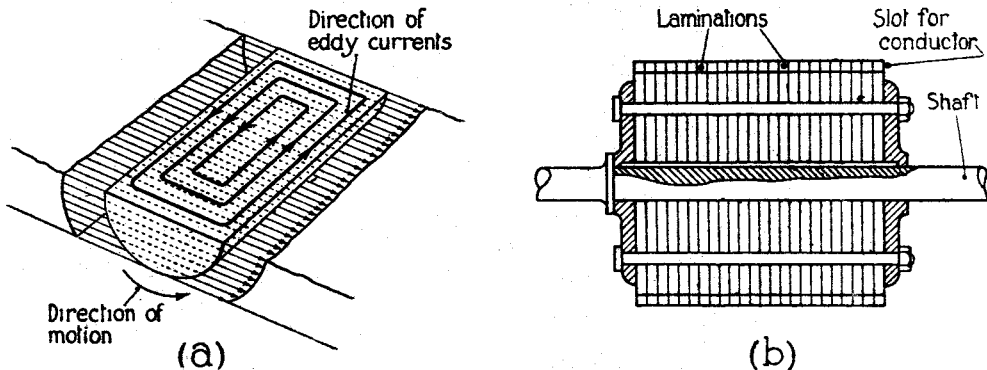


FIG. 8, CHAP. IV.—(a) Eddy currents. (b) Laminated armature.

These are called eddy currents, and represent a wastage of energy in the form of heat, which is doubly undesirable inasmuch as the heating of the core results in heating of the conductors with a consequent increase in resistance. Eddy currents can never be entirely eliminated, but can be reduced considerably by increasing the resistance of the paths in which they flow. The desired end is attained by building up the core from thin circular plates of soft iron called armature stampings, each stamping being insulated from those adjacent by thin paper, varnish, or the natural mill-scale of the plates. This type of construction is known as lamination, and the core is described as a laminated iron core (fig. 8b). The actual conductor may consist of a single loop or a number of complete turns in series with a resulting increase in the amplitude of the generated E.M.F. Of any complete loop or turn of wire in the winding, only those portions which move transversely across the flux are instrumental in generating E.M.F. and these portions are referred to as "inductors". The portions not contributing to the E.M.F. are called end connectors, or coil sides, according to the type of winding. The production of an E.M.F. in a wound armature is illustrated in fig. 9, in which, for clearness, only a single pair of field poles is shown. In the position shown in fig. 9a the inductors (shown by circles) are in the neutral plane and no E.M.F. is being generated. As the armature rotates in a counter-clockwise direction the inductors move in such a manner that a change occurs in the flux linking with each complete turn of the winding. In fig. 9b the armature has moved through 30 electrical degrees, the direction of the E.M.F. is as shown by the conventional signs and its magnitude is $\mathcal{E} \sin 30^\circ$. After rotation through 90 electrical degrees (fig. 9c) the inductors are situated directly under the centres of the field poles; at this instant the flux is undergoing a reversal of direction through the loop and the E.M.F. reaches its maximum value. In the position shown in fig. 9d, i.e. after rotation through 120°,

CHAPTER IV.—PARAS. 13-14

the E.M.F. is still in the same direction, but its magnitude is decreasing, while in fig. 9e the inductors are once more in the neutral plane and the E.M.F. is zero. In fig. 9f the armature has turned through a further 30° , making 210° in all, and the direction of E.M.F. is now reversed, while further rotation, until the inductors are once more in the neutral plane, will complete one cycle of E.M.F.

Slip rings

13. In fig. 9 the external circuit is shown as if it were directly connected to the armature winding, which is not practicable owing to its rotation. In the single-phase alternator each end of the whole winding is connected to one of a pair of slip rings, which are of brass or bronze and are shrunk on to the shaft over mica or micanite insulation (fig. 10). The rings are insulated from each other by mica or fibre washers. The external circuit is then connected to the winding by brushes, which are of hard carbon, graphitic carbon or copper gauze, the material depending to some extent upon the speed of rotation. The brushes are held in light but firm contact with the rings by metal springs.

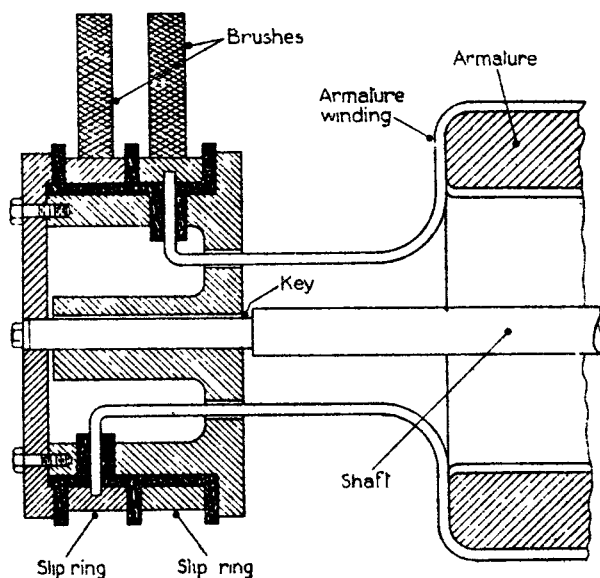
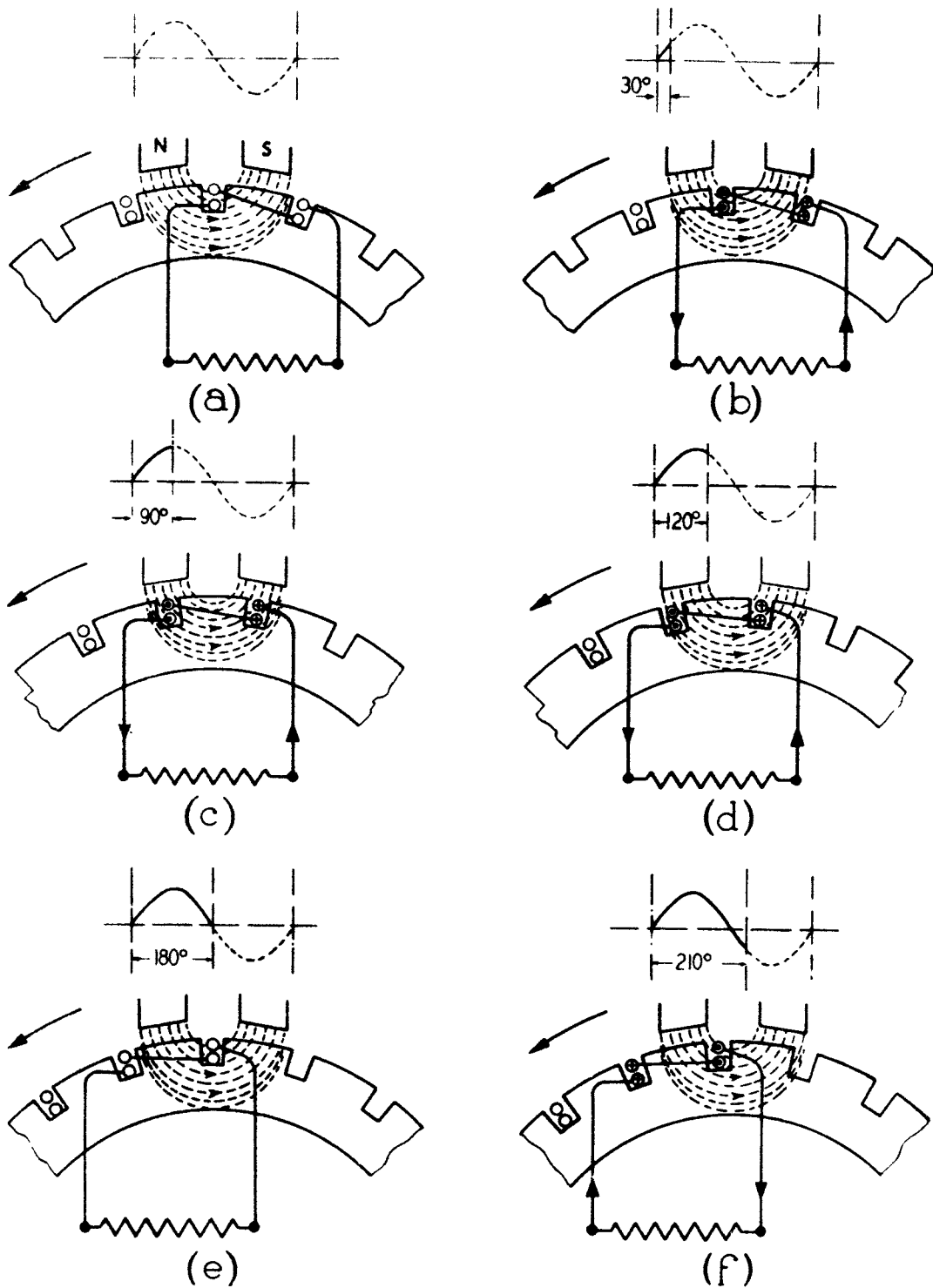


FIG. 10, CHAP. IV.—Slip rings and brushes.

Rotating field type

14. Instead of employing a wound armature a construction called the salient pole type is sometimes adopted. Fig. 11 shows the field and armature windings of an eight-pole machine of this type in diagrammatic form. If the stationary winding $W_1, W_2 \dots W_8$ were energized by a direct current an alternating E.M.F. would be induced in the winding carried by the rotating system. If, however, the arrangement is as shown in the diagram, the rotating winding being energized by a direct current, an alternating E.M.F. will be induced in the stationary winding. In order to avoid confusion the term "rotor" is used for the moving member and "stator" for that which is stationary. A salient pole machine may be designed either with a rotor armature and stator field, or with stator armature and rotor field as in the diagram; practically all high voltage alternators of modern design are of the rotor field type, for the mounting, ventilation and insulation of the windings are then much facilitated, and the difficulties of highly insulating the slip rings and brushes are avoided. The constant P.D. between the slip rings rarely exceeds 200 volts, although the alternating P.D. between the ends of the stator winding may be of the order of thousands of volts. In fig. 11 the stator coils are denoted by $W_1, W_2 \dots W_8$ and the



E.M.F. IN ARMATURE OF ALTERNATOR

rotor windings and their poles by N, S, N, S. The armature coils are placed end-on to the poles in the diagram so that their inter-connection may be clearly seen, but actually the coils face the pole pieces. Taking any one coil, e.g. W_1 , the inductors are those portions marked k, k . The E.M.F. induced in those inductors which are being transversely by N-poles will be in the reverse direction to that induced in the inductors which are being transversely by S-poles, and the coils must be connected alternately right-handed and left-handed as shown, so that the total effective E.M.F. shall be additive.

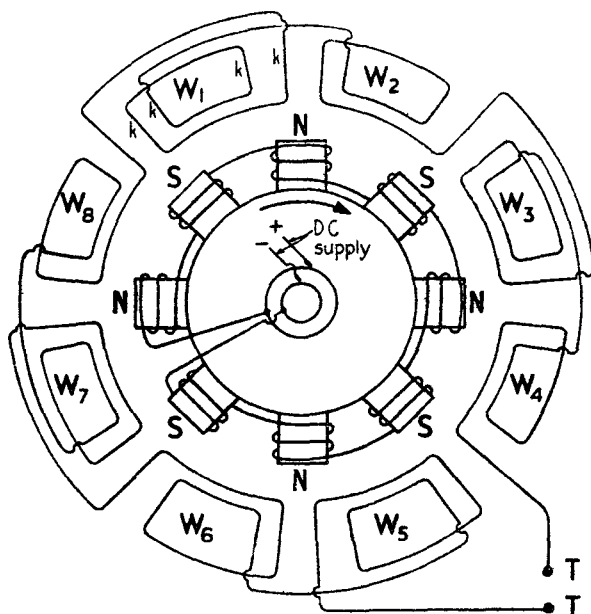


FIG. 11. CHAP. IV.—Alternator windings—stator armature and rotor field.

Direct current generators

15. The direct current generator makes use of exactly the same principle as the simple alternator described in paragraph 3, the machine being invariably designed with stator field and rotor armature. An alternating E.M.F. is produced in the armature conductors, but the current in the external circuit is uni-directional owing to the action of a device known as a commutator. Let us consider the loop which has already been assumed to rotate in a uniform magnetic field, and in which an alternating E.M.F. is being generated. If instead of being brought to slip rings the two ends of the rotating loop are connected to a single ring which is split longitudinally, i.e. in the direction of the axis of rotation, at two points diametrically opposite, the direction of the current in the external circuit will be reversed when the armature rotates through 180° . Referring to fig. 12a, at the instant depicted, the inductors, P and Q are cutting the flux at an angle of 90° and maximum E.M.F. is induced at this instant. By the right-hand rule, the E.M.F. in the inductor P is acting away from and that in the inductor Q towards the reader. Inductor P is connected to the segment (b) of the split ring, while Q is connected to the segment marked (a). As a result, the combined E.M.F. of P and Q causes a current to flow in the external circuit in the direction shown by the arrows. After rotation through 90° , neither P nor Q are cutting the flux and the current falls to zero. During the next 90° of rotation the inductor P is moving toward the position occupied in the diagram by Q, while inductor Q is moving toward that occupied by P, so that the E.M.F. in the inductor P is towards, and that in the inductor Q is away from the reader. The brushes are, however, bearing on the opposite segments and the current in the external circuit is in the same direction as before, thus, although the current in the inductors

CHAPTER IV.—PARA. 16

themselves is constantly reversing in direction, the current in the external circuit is uni-directional. In effect the external current consists of half sine waves, flowing always in the same direction but varying in intensity from zero, through a maximum value, and back to zero again, two repetitions of this series of instantaneous values occurring during one rotation of 360 electrical degrees, fig. 12b. The split ring device causing the reversal of current in the external circuit is called a simple or two-part commutator, each of the conducting portions being termed a commutator segment.

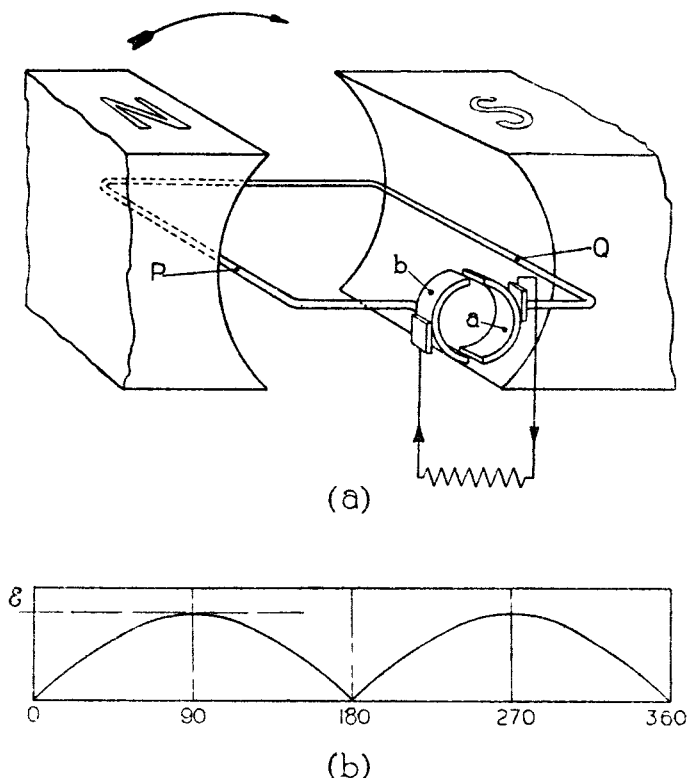


FIG. 12, CHAP. IV.—Principle of commutation.

16. In order to obtain a current which is to all intents and purposes of uniform intensity, it is necessary to use a large number of inductors spaced round the armature core and suitably inter-connected. In fig. 13a two loops are mounted at right angles upon an axis of rotation (not shown) and the ends brought to a commutator having four segments. At the instant depicted the inductors P and Q are cutting the flux at an angle of 90° , and the peak E.M.F., $\mathcal{E}$ volts, is induced in each, while the inductors R and S are not cutting the flux and no E.M.F. is induced in them. During the ensuing rotation through an angle of 45° , the E.M.F. in the inductors P, Q, will diminish and that in the inductors R S will increase, and when this amount of rotation has taken place the E.M.F. induced in all four inductors will be the same, viz. $\mathcal{E} \sin 45^\circ = 0.707 \mathcal{E}$. At this instant the brushes, which were previously bearing on the commutator segments c and d, momentarily connect segment (d) to segment (b), and segment (c) to segment (a). During the next 90° of rotation they will bear on the segments (b) and (a) only. The E.M.F. during this portion of the rotation will be that induced in the loop R S, which is increasing in value from $0.707 \mathcal{E}$ to $\mathcal{E}$, reaching the latter value after rotation through 90° from its original position.

During one complete revolution, therefore, the E.M.F. applied to the external circuit will vary as shown by the heavy line in fig. 13b, and never rises above the value $\mathcal{E}$ volts nor falls below the value $0.707 \mathcal{E}$ volts.

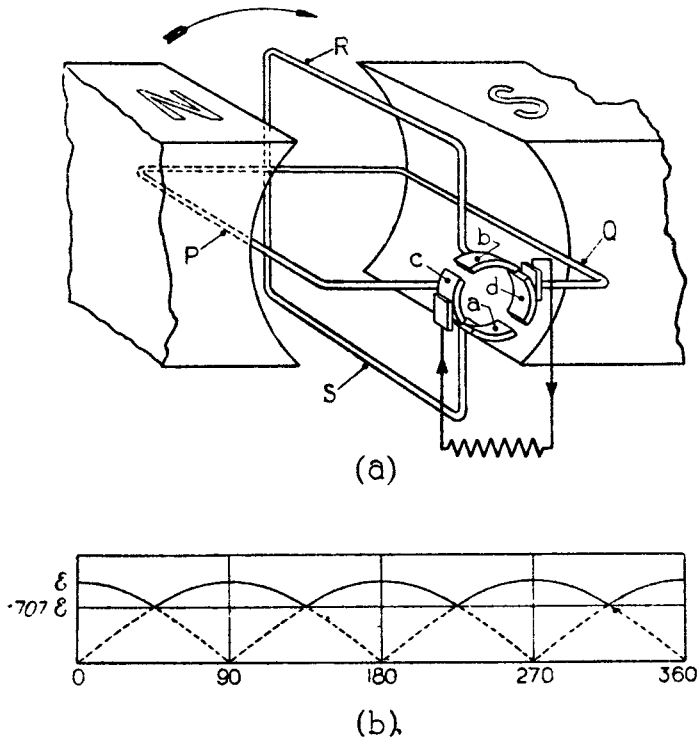


FIG. 13, CHAP. IV.—Approach to uniform E.M.F.

17. It will be observed that with this arrangement the effective E.M.F. is only that of the loop to which the external circuit is connected by the commutator and brushes, the E.M.F. in the other loop being applied to no useful purpose. By suitably interconnecting the loops the E.M.F. in both loops may be utilized. The principle of the closed winding, as it is called, may be explained with reference to fig. 14, in which the armature has two coils each having two complete turns, and the winding is closed upon itself. That this is so is easily traced out, starting say, from segment (a) via inductors PQ to segment (c), then via inductors RS to segment (b). From this point the winding continues via the inductors $Q'P'$, to segment (d) and inductors $S'R'$, terminating upon segment (a) and thus closing the winding. Assuming that the brushes are so thin that they just bridge over two adjacent segments at the instant of commutation, it will be seen that the E.M.F. applied to the external circuit varies from $\mathcal{E}$ to $1.414 \mathcal{E}$ during the rotation of the armature. Let the brushes—which have been omitted from the diagram for clearness—be on the neutral axis, i.e. midway between the poles, so that at the instant depicted in the diagram they are bearing on segments (a) and (b). From the segment (a) we may now trace two paths through the armature

(1) (a) PQ (c) RS (b)

In this path the indicators R and S are cutting the flux at maximum velocity, while P and Q are not cutting flux hence the general E.M.F. in this path is $\mathcal{E}$ volts.

(2) (a) $R'S'$ (d) $P'Q'$ (b)

CHAPTER IV.—PARA. 18

This path is in parallel with path (1) and also has a generated E.M.F. of $\mathcal{E}$ volts. When the armature has rotated through very nearly 45° , the upper and lower brushes are just about to leave the segments (a) and (b) respectively. All inductors are now cutting flux at approximately 45° , and the E.M.F. induced in each pair of inductors is $0.707 \mathcal{E}$ volts. Tracing out the paths as above, then, it is seen that in each path the generated E.M.F. is $2 \times 0.707 \mathcal{E} = 1.414 \mathcal{E}$. Thus, during rotation, the generated E.M.F. varies as shown in fig. 14b. The dash line shows the E.M.F. of one loop and the chain line that of the other, while the solid line is the sum of the two. The fact that there are two paths in parallel through the armature implies that the internal resistance of the winding is reduced. The above arrangement is however hardly a practicable arrangement because commutation troubles would arise.

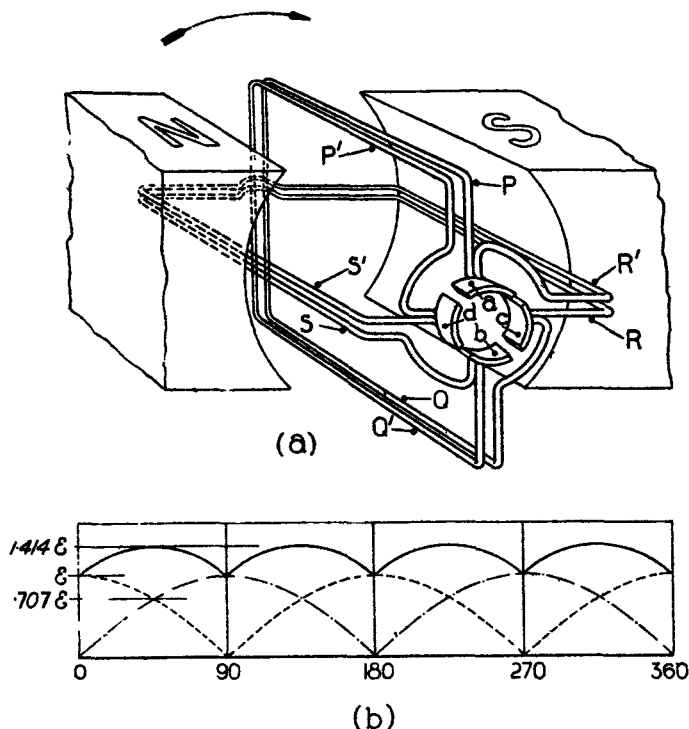


FIG. 14, CHAP. IV.—Principle of closed winding.

18. The practical form of D.C. generator consists of the following essentials:—

- (i) The field magnets, which provide the magnetic field in which the conductors rotate.
- (ii) The field magnet windings, which are conductors carrying the magnetizing current for the field magnets. These windings are of course absent in machines fitted with permanent field magnets.
- (iii) The armature core and spider. The core is of soft iron and is laminated as in the alternator, for the same purpose. The spider is an arrangement by which the armature core is mounted on its shaft.
- (iv) The armature winding, which is the system of conductors in which the E.M.F. is induced. Various forms of winding are described below.
- (v) The commutator.

(vi) The brushes, which are metal or carbon current collectors. These rest upon the commutator and convey the current to the external circuit. The brushes are supported by the brush gear, which allows the position of the brushes to be shifted round the periphery of the commutator for reasons which will be apparent later.

19. Except in small low power machines, an electro-magnetic field system is generally adopted, the disadvantage of permanent magnets being as already stated when dealing with the alternator. In spite of this disadvantage, certain low power generators for W/T purposes in aircraft utilize a permanent magnet field in order to remove the necessity for providing a heavy magnetizing current. Small D.C. generators are usually bi-polar, having one N-pole and one S-pole, but for outputs greater than about five kilowatts, multi-polar construction is practically universal, either a four-pole or six-pole design being adopted; turbo-generators running at very high speed generally utilize the bi-polar construction even for outputs of 1,000 kilowatts or more. The field magnet windings are generally supplied with current by the machine itself, and the latter is then said to be self-excited, differing in this respect from the alternator, which is usually provided with field magnet excitation from a separate source. It must be observed that if the field magnet is absolutely without magnetic properties when the armature is initially set in rotation, no E.M.F. can be developed in the latter and consequently it can supply no excitation to the field. Usually, however, the field magnets possess some slight residual magnetism, and on rotation the armature generates a small E.M.F., consequently a small current is established in the field windings. This current is normally in such a direction as to strengthen the residual magnetism of the field magnet so that the generated E.M.F. is progressively increased, and is commonly said to "build up" to the final value. If the residual magnetism is destroyed, e.g. if the pole pieces are removed from the carcass during the course of repairs, the machine may fail to excite on first running after re-assembly. An initial field current must then be provided by means of an accumulator battery. The commutator consists of a number of segments of hard copper, usually manufactured by drop-forging. These are insulated from each other by sheet mica, and from the frame with which they are secured to the shaft by micanite rings. The stationary brushes rest on the segments and carry the current to and from the armature winding.

20. The armature coils or bars are connected to each other and to the commutator in one of two methods known as wave winding and lap winding respectively. In both forms, two separate inductors are placed in each slot in the armature. A single winding unit may be defined as a portion of the winding which is connected between two commutator segments, and the pitch of the winding as the distance between two inductors which are in direct connection. Instead of actually stating the pitch in terms of length it is more convenient to express it as the number of inductors passed over in so inter-connecting any two. The inductors forming the active portions of a winding unit must be situated in fields of opposite polarity in order that the E.M.F.'s induced in each shall be additive. The width of a winding unit will therefore be, as nearly as possible, equal to the pole pitch, i.e. the circumference of the armature divided by the number of poles. The inter-connection at the commutator end is referred to as the front pitch, y_f , and that at the opposite end as the rear pitch, y_r . For example, in fig. 15, the inductors are numbered consecutively, and the inductor (1) is connected at the rear end to the inductor (6); the rear pitch is therefore 5. At the front end the inductor (6) is joined to the inductor (11), and the front pitch is also 5. If the front and rear pitches differ, the difference is always an even number and

the mean pitch, y , is equal to $\frac{y_f + y_r}{2}$. The front and rear pitches are so adjusted that the

winding shall be closed upon itself. Starting from any inductor, and tracing through the whole system of winding, each inductor must be traversed once and once only, before returning to the original inductor. The winding is then said to be re-entrant.

CHAPTER IV.—PARAS. 21-22

Wave winding

21. In this type of winding both front and rear pitches are invariably odd numbers. If there are N inductors, the mean pitch is $(N + 2)$, or $(N - 2)$, divided by the number of field poles. If the mean pitch is odd, the front and rear pitches are equal, i.e. $y_f = y_r = y$. If the mean pitch is even, the front pitch may be greater than the mean pitch by unity, and the rear pitch less than the mean pitch by unity, or

$$y_f = y + 1$$

$$y_r = y - 1.$$

Alternatively, it is permissible to arrange that $y_r = y + 1$ and $y_f = y - 1$. Fig. 15 is what is called an expanded winding diagram and illustrates the method of wave winding in a four-pole machine. The armature and pole pieces are imagined to be laid out on a plane, the inductors being numbered (1) to (22) and the commutator segments (a) to (k). As $N = 22$, the mean pitch may be either $\frac{22 + 2}{4} = 6$ or $\frac{22 - 2}{4} = 5$. If the value 6 were taken, the front and rear pitches would be 7 and 5 respectively. In the diagram, however, the mean pitch has been taken as 5, and the front and rear pitches are both equal to the mean pitch. Commencing from the commutator segment (a) two complete winding units are shown in heavy line to assist in tracing the circuit. The position of the poles is also indicated, but it must be remembered that they are really situated above or below the paper. Assuming they are above, the direction of the induced E.M.F. is found by the right-hand rule, and is marked on the inductors by arrows. As the inductors shown by dotted lines lie underneath those immediately adjacent (solid lines), certain inductors must be given an E.M.F. although, as drawn, they appear to lie outside the influence of the field. For example, inductor (5) lies beneath inductor (4), and equal E.M.F.'s are induced in each.

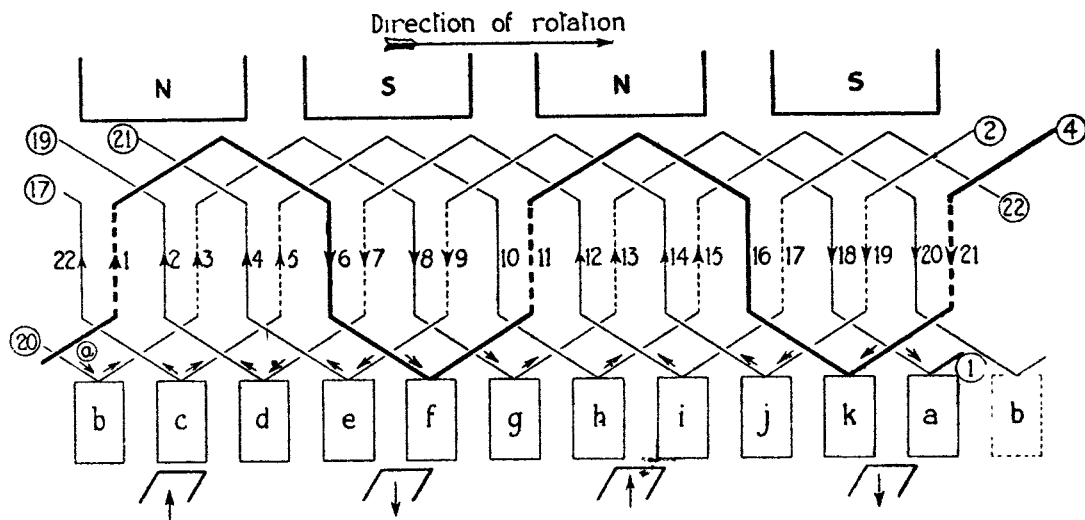


FIG. 15, CHAP. IV.—Principle of wave winding.

22. The positions on the commutator at which the brushes should be placed are found as follows. Assuming that in the front connectors current will only flow as a result of the E.M.F. in the inductors connected thereto, place arrows beside the connectors to indicate the direction of current. It will then be seen that current flows both toward and away from certain segments, while at others current flows only in a single direction, and these are the positions at which brushes should be placed. At segment (c) current flows away from the commutator to inductors (22) and (5); current may therefore enter the winding by a brush placed on this segment. At

segments (k) and (a) current flows only from inductor (21), inductor (16) having no induced E.M.F. and a second brush placed between these segments will convey current away from the commutator. These brushes are 90° apart and are sufficient for the correct performance of the machine. Alternatively, the brushes could be located on segments (f) and between segments (h) and (i). In practice, therefore, two pairs of brushes are often fitted and connected in parallel, in order to reduce the current density in the commutator segments.

Lap winding

23. In a lap winding the front and rear pitches are both odd, and differ by 2, making the mean pitch an even number. Fig. 16 shows the developed diagram of a four-pole machine having 24 inductors and 12 commutator segments. The mean pitch in this form of winding is given by

$$y = \frac{\text{number of inductors}}{\text{number of poles}},$$

and in the given example is 6. The front pitch is $y - 1 = 5$, and is contrary to the direction of rotation, hence it is given a negative sign and becomes -5 . The rear pitch is in the direction of rotation and is equal to $y + 1$, or 7. One complete winding unit, shown in heavy line, starts from segment (a), thence to inductor (1), via the rear connector to inductor (8) and by the front connector to segment (b). The winding then continues in an overlapping manner from which the term lap winding is derived. Arrows are placed upon the inductors to indicate the direction

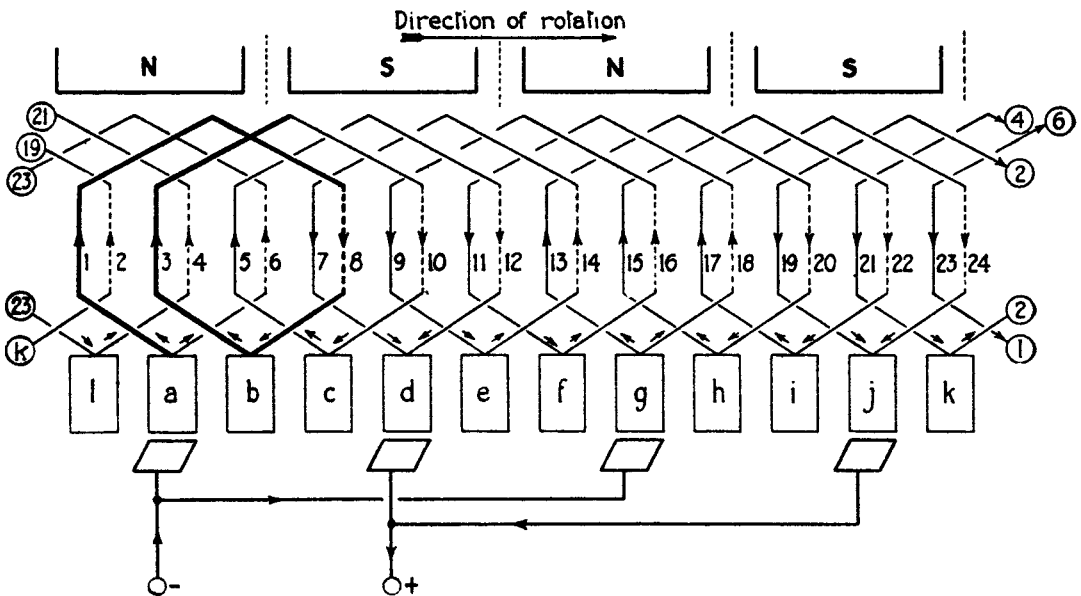


FIG. 16, CHAP. IV.—Principle of lap winding.

of the induced E.M.F., and beside the front connectors to denote the direction of current. It will be seen that current reaches the commutator from inductors (7) and (12) at segment (d), and from inductors (19) and (24) at segment (j), while current leaves the commutator at segments (a) and (g). It is possible to trace four separate paths through the armature, viz :—

- (i) from segment (a) through inductors (1, 8, 3, 10, 5, 12) to segment (d) ;
- (ii) from segment (a) through inductors (6, 23, 4, 21, 2, 19) to segment (j) ;
- (iii) from segment (g) through inductors (13, 20, 15, 22, 17, 24) to segment (j) ;
- (iv) from segment (g) through inductors (18, 11, 16, 9, 14, 7) to segment (d).

CHAPTER IV.—PARAS. 24-25

Thus there are as many paths through the armature as there are poles. Current is taken from the armature by brushes placed on the commutator at segments (d) and (j), and returns to it by brushes placed at segments (a) and (g). In a lap wound machine, therefore, one pair of brushes must be provided for each pair of poles.

24. In the foregoing explanation, the armature is assumed to consist of a series of single inductors suitably connected to each other and to the commutator, but in practice each winding unit may consist of a coil of several turns, particularly in machines designed to give a high voltage. Thus in fig. 16, taking the winding unit shown in heavy line and starting from segment (a), the conductor would follow the path through the slot occupied by inductor (1), via the back connection occupied by inductor (8), and then, instead of terminating upon the segment (b), would be carried on via a front connection to the slot occupied by inductor (1) thence to that occupied by inductor (8) and so on, forming a coil of several complete turns which is finally terminated on the commutator at segment (b). This is termed an armature coil, and is generally made by winding double cotton-covered wire on a suitably shaped wooden former, the turns being bound together with cotton tape and varnished.

Magnitude of E.M.F.

25. The total E.M.F. developed in the armature is easily found by the following method.

Let

n = r.p.m.

Φ = total flux of per pole.

p = number of pairs of poles.

a = number of pairs of paths through armature.

N = number of armature conductors.

When an armature loop, i.e. a pair of conductors, moves from a position in which it embraces the whole of the flux from a N-pole, to a similar position with reference to a S-pole, the total flux change is 2Φ , because each tube of flux is cut twice during the movement, which takes place in $\frac{1}{p}$ of a revolution. The time of one revolution is $\frac{1}{n}$ of a minute or $\frac{60}{n}$ of a second; the time

taken by the above change of flux is therefore $\frac{1}{2p} \times \frac{60}{n}$ second. The average E.M.F. E_1 induced in one loop is equal to the total change of flux divided by the time occupied by the change and therefore

$$E_1 = \frac{2\Phi}{\frac{1}{2p} \times \frac{60}{n} \times 10^8} \text{ volts.}$$

The number of armature loops in series is $\frac{N}{4a}$ and the total average E.M.F. is

$$\begin{aligned} E &= \frac{4n p \Phi}{60 \times 10^8} \times \frac{N}{4a} \text{ volts} \\ &= \frac{n N \Phi}{60 \times 10^8} \times \frac{p}{a} \text{ volts.} \end{aligned}$$

For a wave wound armature, $a = 1$, and

$$E = \frac{n N \Phi}{60 \times 10^8} \times p \text{ volts.}$$

For a lap wound armature, $a = p$, and

$$E = \frac{n N \Phi}{60 \times 10^8} \text{ volts.}$$

Example

The flux in an eight-pole dynamo is 2,400,000 lines per pole. The armature has 740 conductors and is lap-wound. Find the induced E.M.F. at a speed of 800 r.p.m.

$$\Phi = 2.4 \times 10^6$$

$$n = 800$$

$$N = 740$$

$$E = \frac{800 \times 740 \times 2.4 \times 10^6}{60 \times 10^3}$$

$$= 236.8 \text{ volts.}$$

Power output of D.C. generator

26. Up to the present, reference has been made to E.M.F. generated by the rotation of conductors in the magnetic field and to current flowing in various portions of the circuit, without any reference to energy converted from mechanical into electrical form. When the external circuit is incomplete, no electrical energy is supplied to the external circuit, and the energy supplied by the steam engine, petrol motor, or other means of rotation, has only to turn the armature against the various forms of friction which are present. When the external circuit is completed, and a current is established, electrical energy is supplied to the external circuit, the rate at which this energy is supplied being called the power output of the generator. This power must be supplied by the mechanical source, and the load on the latter increases with the power output. The current flowing through the external circuit must also flow through the armature, and the latter then constitutes a system of current carrying conductors situated in a magnetic field. Now in the preceding chapters it has been shown that such a conductor experiences a force, which tends to cause it to move in a direction which may be found in the left-hand rule. Applying this to the simplest instance, i.e. a single rotating loop carrying current in a bi-polar field, it is found that the force tends to turn the loop in the opposite direction to that in which it is actually rotating, and must be overcome by the supply of additional power by the mechanical source.

Torque

27. The turning moment just referred to is called the electro-dynamic torque of the armature and is measured in pounds-feet, because it is the product of the force applied (in pounds) and the radius (in feet) of the path in which each conductor tends to turn. If a torque of T pounds-feet is allowed to act through one complete revolution, the work done is $2\pi T$ foot-pounds. Alternatively, if the radius of a certain path is r feet, and a force of P pounds is acting in this path, the work done in one rotation is $2\pi rP$ foot-pounds, the product $rP = T$ being the torque exerted. If rotation takes place at a speed of n r.p.m., work is performed at a rate of $2\pi nT$ foot-pounds per minute or $\frac{2\pi nT}{60}$ foot-pounds per second, and this is the power required to cause the rotation. It is convenient to convert this into horse-power (H.P.) since this unit of power is generally used for mechanical power. As one H.P. = 33,000 foot-pounds per minute or 550 foot-pounds per second, the power which the mechanical source must supply, in order to overcome the electrodynamic torque, is $\frac{2\pi nT}{33,000}$ H.P. This mechanical power is completely converted into electrical power, the latter being EI_A watts, where E is the E.M.F. generated and I_A the armature current. As EI_A watts are equal to $\frac{EI_A}{746}$ H.P.,

$$\frac{2\pi nT}{33,000} = \frac{EI_A}{746}$$

$$T = \frac{EI_A \times 33,000}{746 \times 2\pi n}$$

CHAPTER IV.—PARAS. 28-29

In a previous paragraph it was shown that for either lap or wave wound armatures, the E.M.F. generated is directly proportional to the magnetic flux per pole, ϕ , and directly proportional to the speed n . As a result the torque becomes independent of the speed and dependent only upon the amount of flux per pole and the armature current, or

$$T \propto \phi I_A.$$

Armature reaction

28. Hitherto, the magnetic field in which the armature is rotated has been considered to be of uniform density, and established entirely by the field magnets. As we have seen, the armature itself is a system of current-carrying conductors, and must also set up a magnetic field in the space between the field poles. The total flux in this space is therefore made up by the superposition of the armature field upon the field due to the stator winding. Before proceeding further, it is pointed out that in diagrams illustrating the results of this superposition, it is customary to omit the commutator, and to show the brushes as bearing directly upon the armature conductors. This is not unreasonable, for the commutator may be regarded merely as a portion of the armature winding which has been bared of insulation in order to make contact with the external circuit by means of the brushes. Fig. 17a shows the field magnets and armature in section, the direction of current in the armature conductors being indicated conventionally. The armature current is here supposed to set up no magnetic flux, and in these circumstances the proper position of the brushes would be as shown, i.e. making contact with the conductor in which no E.M.F. is being generated at the moment.

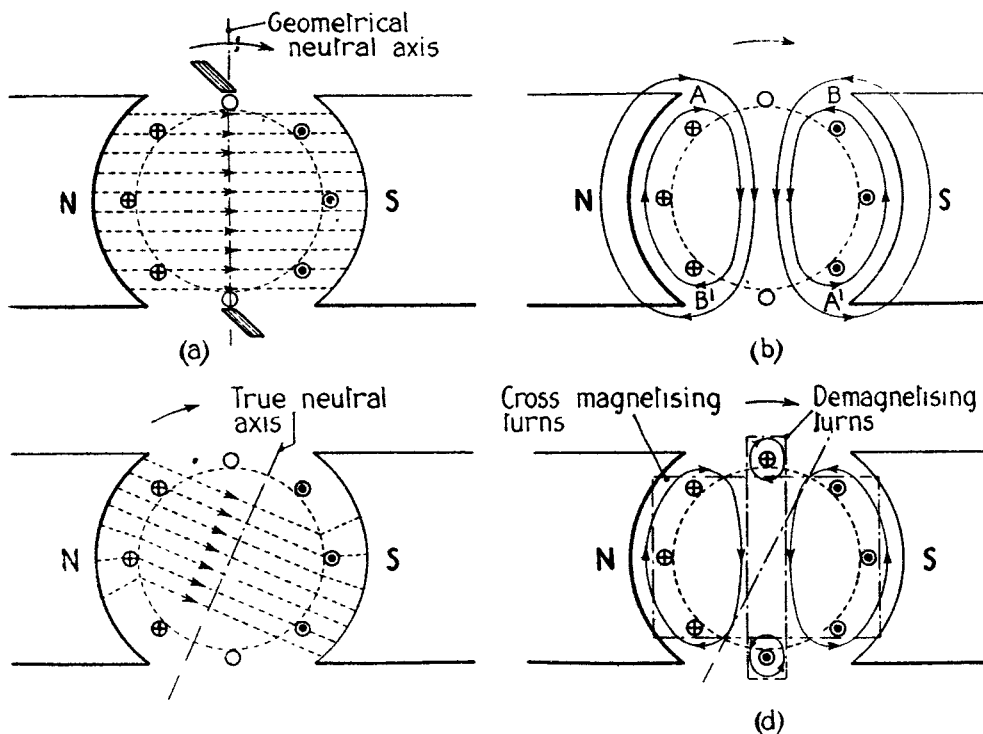


FIG. 17, CHAP. IV.—Armature reaction in D.C. generator.

29. The field due to the current in the armature is shown in fig. 17b. It is seen to be in the same direction as the stator field at the points A A', and in the opposite direction to the stator field at the points B B'. With reference to the direction of rotation, the points A A' are referred

to as the trailing pole tips, and the points $B B'$ as the "leading pole tips". The direction of the armature field being at right angles to the stator field, this effect is referred to as cross-magnetization, and its result is to distort the total effective flux in the direction of rotation as shown in fig. 17c. If the brushes are to be placed upon the conductors in which no E.M.F. is being induced, they must be shifted round in the direction of rotation, so that their position is upon the electrical neutral axis as indicated in this diagram. The current distribution in the armature conductors now differs from that shown in fig. 17a, becoming as illustrated in fig. 17d. The conductors on the armature may now be divided into two groups at right angles to each other, one group causing cross-magnetization as before, and the other causing a magnetic field which is parallel to the stator field but of opposite polarity, and therefore this component tends to weaken the flux. This weakening is referred to as the demagnetizing effect. The principal results of this factor in armature design are (i) to cause a reduction in E.M.F. for a given field current, and (ii) to cause an uneven distribution of the voltage between the various armature coils, and consequently a greater strain on the insulation between some commutator segments than between others. The commutator must therefore be designed with a larger margin of safety than would otherwise be necessary.

Sparking

30. Sparking at the brushes is caused by the inductance of the armature windings, which gives rise to what is called the reactance voltage of the coil. This is a true E.M.F. of self-induction due to the change of current in the coil, and must not be confused with the E.M.F. induced in it owing to its rotation. Fig. 18a shows a portion of an armature winding and its connection to the commutator, in the vicinity of the positive brush. The coil B' is short-circuited by the brush; current is flowing towards the brush from c' to c , i.e. downward, in the coil C' , and from a' to b , i.e. downward, in the coil A . In fig. 18b the coil B is about to undergo commutation.

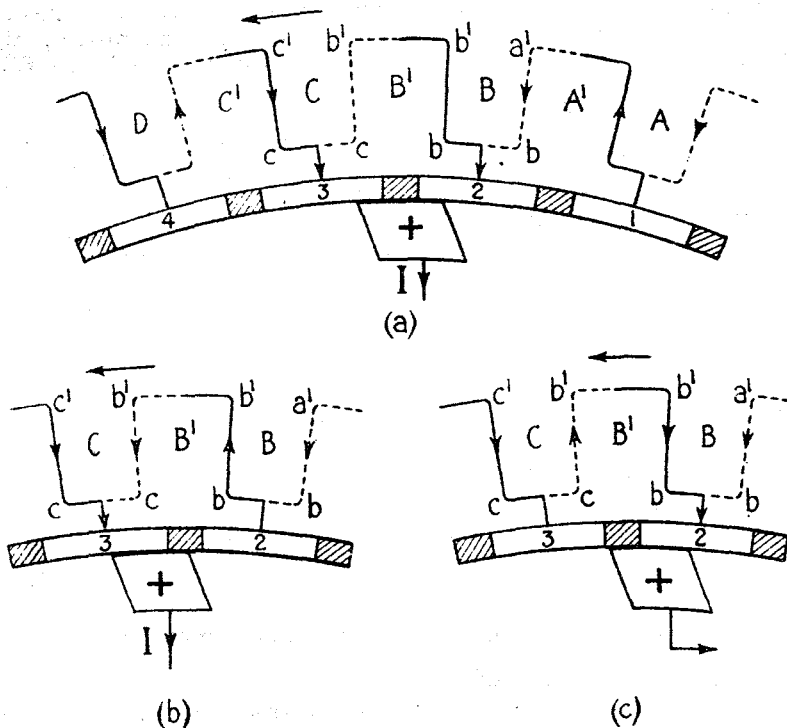


FIG. 18, CHAP. IV.—Commutation.

CHAPTER IV.—PARAS. 31-32

current in the conductor $b b'$ is in the upward direction, while in fig 18c, the commutation has been performed and the current in $b b'$ is in the downward direction. The total armature current is I , and in the armature itself there are two paths in parallel, so that commutation implies a change of current from $+\frac{1}{2}I$ to $-\frac{1}{2}I$, or a change of I amperes during the time of commutation, i.e. while the brush is short-circuiting segments (2) and (3), as in fig. 18a. Under ideal conditions, the change of current would take place uniformly, as shown by the dotted line BC in fig. 19a. Here OA is the current $+\frac{1}{2}I$ before commutation, and NC the current $-\frac{1}{2}I$ afterwards. In practice, however, the reversal is delayed owing to the counter-E.M.F. of self-induction due to the change of current in the coil, and the actual change of current may be as shown by the curved line BE. At the end of the short-circuit time, the current has assumed the value NE instead of NC and must change from NE to NC very suddenly. It is this sudden change of current which causes a large induced E.M.F. and a corresponding spark to pass from the segment (3) to the brush.

Methods of reducing sparking

31. One method of reducing sparking is by giving the brushes a forward lead in order to bring the short-circuited coil into a reversing field. It has already been stated that one effect of armature reaction is to necessitate a forward lead of the brushes so that commutation may take place in zero field. By moving the brushes still further in the same direction the coil is commutated in a reversing field and remains short-circuited after it has commenced to cut the flux in the new direction. The induced E.M.F. due to this cutting of flux opposes the original current in the coil. If an excessive lead is given, the induced E.M.F. will more than balance the reactance E.M.F. and the current in the coil will be greater than $\frac{1}{2}I$ in magnitude so that at the end of the short-circuit time it will have to change abruptly and sparking will again occur (fig 19b). When the induced and reactance voltages are exactly balanced, at the end of the commutation period, the change of current is somewhat as shown in fig. 19c and no sparking takes place. The use of carbon brushes is of great assistance in the attainment of sparkless commutation. The resistance of the carbon-copper contact is comparatively high, and this tends to cause the current change during commutation to follow the straight line BC of fig. 19a.

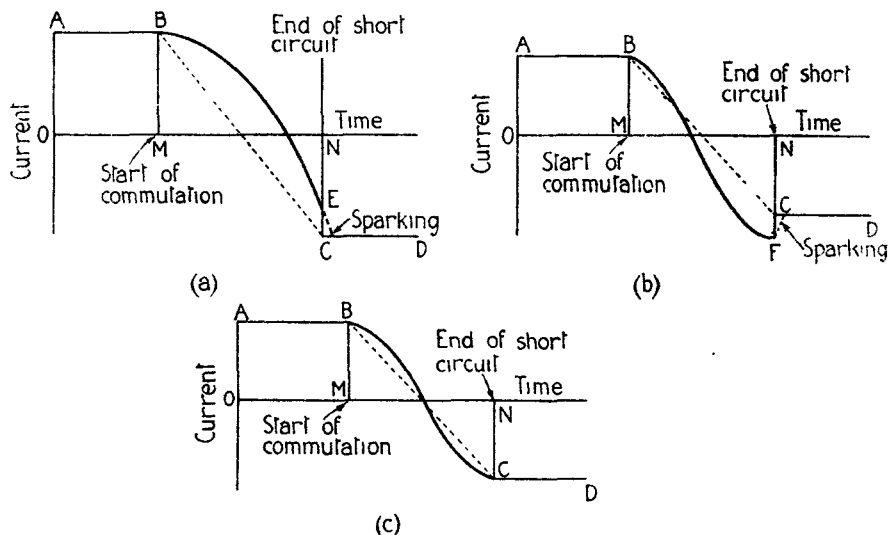


FIG. 19, CHAP. IV.—Current variation during commutation.

32. Minor advantages of carbon brushes are:—

- (i) If properly "bedded" in the first place, carbon brushes lubricate and polish the commutator.

(ii) If sparking does occur, less damage to the commutator ensues than with copper brushes.

Their chief disadvantages are:—

(iii) Owing to their high resistance (which is essential to obtain the benefit of improved commutation) an IR drop of about 2 volts takes place at each brush. This effect is only serious on low voltage machines.

(iv) The heat generated in the brush to this high resistance raises the temperature of the commutator, and the latter must be made larger than if copper brushes were used, in order to radiate the heat.

In modern machines, inter-poles are often fitted to assist commutation. In such circumstances the brushes are placed upon the electrical neutral axis, and a reversing field is provided by means of a field winding upon special poles placed midway between the main poles. This winding generally carries the full armature current and is therefore a series winding. The polarity must be that of the next main pole in the direction of rotation.

Methods of connecting the field windings

33. Self-excited generators may be classed as series, shunt, or compound wound, the distinction being in the manner in which the field circuit is connected with reference to the armature. The series wound machine is shown diagrammatically in fig. 20a. The field winding

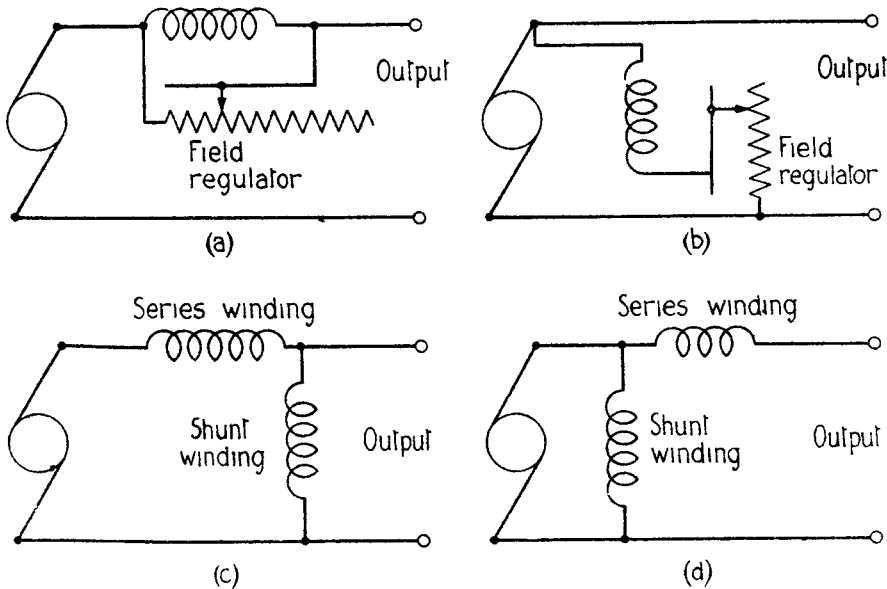


FIG. 20, CHAP. IV.—Methods of connecting field windings.

carries the full armature current or a large fraction thereof, hence to provide the required ampere-turns only a few turns of wire of large cross-section are required. Fig. 20b shows the shunt-wound machine, in which a small portion of the armature current is diverted from the external circuit for the purpose of providing the field ampere-turns, and the number of turns must be large, although the cross-section of the wire is small. The excitation is adjusted by means of a field regulator, which is a rheostat of suitable current-carrying capacity. In the series-wound machines the field regulator is in parallel with the field magnet winding and so diverts a portion of the load current from that path, while in the shunt-wound machine the field regulator is placed in series with the field magnet windings. The compound winding is a combination of the two and may be either "long shunt" or "short shunt" as shown in fig. 20c and fig. 20d respectively. The behaviour of generators with different types of field excitation can be shown

CHAPTER IV.—PARAS. 34-35

by means of the characteristic curves of the generator. These are graphs of the relation between current and voltage when the generator is run at constant speed. The most important of these curves is the "external characteristic" which shows relation between the terminal P.D. and the current flowing in the external circuit.

Separately excited or permanent-magnet field generator

34. As a basis of comparison, it is convenient to use the external characteristic of a separately excited machine, which is identical with that of a machine having a permanent-magnet field, because the stator flux is quite independent of the armature current, although of course the total field is influenced by the effects of armature reaction. In fig. 21 the E.M.F. generated at constant speed is shown by the line AB and is nearly constant for all values of output current. The dotted horizontal line is inserted in order to show the extent of the deviation. The terminal P.D. is shown by the line AC; it falls off gradually as the current output is increased, that is, as the resistance of the external circuit is decreased. The difference between the E.M.F. on no load and the terminal P.D. at full load is called the regulation of the machine, and if the fall in terminal voltage at full load is only slight, the machine is said to have good regulation. The difference between the E.M.F. and the terminal P.D. is chiefly caused by the increase in the potential drop in the armature itself. This is equal to the product of the armature current and the armature resistance. The lower the resistance of the armature, therefore, the better will be the regulation of the generator. The small potential drop at the brushes must also be taken into account.

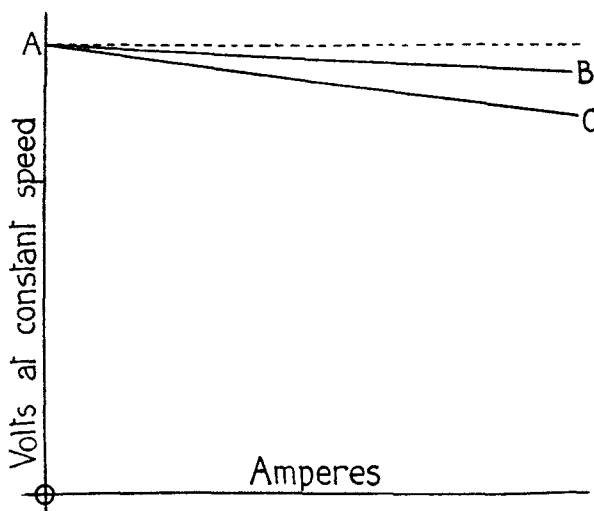


FIG. 21, CHAP. IV.—External characteristic of separately excited generator.

The shunt-wound generator

35. The external characteristic of this type of generator is given in fig. 22. In the first place, consider the E.M.F. generated. If no current is taken by the external circuit, the E.M.F. builds up to some steady value, and as the machine has only to supply the small field magnetizing current, the terminal P.D. is practically equal to the E.M.F. OA. If an external circuit of variable resistance is connected to the generator terminals, the variation of terminal P.D., within certain limits, is similar to that of a separately excited machine, as shown by the portion AB of the characteristic. For a given load current, the fall of P.D. is somewhat greater because as the terminal P.D. falls, the field current diminishes and the excitation is correspondingly reduced. If the external resistance is reduced below the value corresponding to the point B, however, the terminal P.D. falls to some very small value such as OC, which is maintained almost entirely

by the residual magnetism. For any given machine, then, there is some maximum current output which cannot be exceeded. The corresponding value of external resistance is found by drawing a tangent OTP to the curve from the origin O. This tangent should just touch the characteristic curve at the point T. (In the diagram a slight clearance has been left for distinctness.)

The slope $\frac{PN}{ON}$ has the dimensions $\frac{\text{volts}}{\text{amperes}}$ and is therefore a resistance. If the resistance of the external circuit falls below this value, the machine fails to excite. It is obviously desirable to start up a shunt-wound generator on open circuit, and allow the E.M.F. to build up to its normal value OA, before connecting the external circuit.

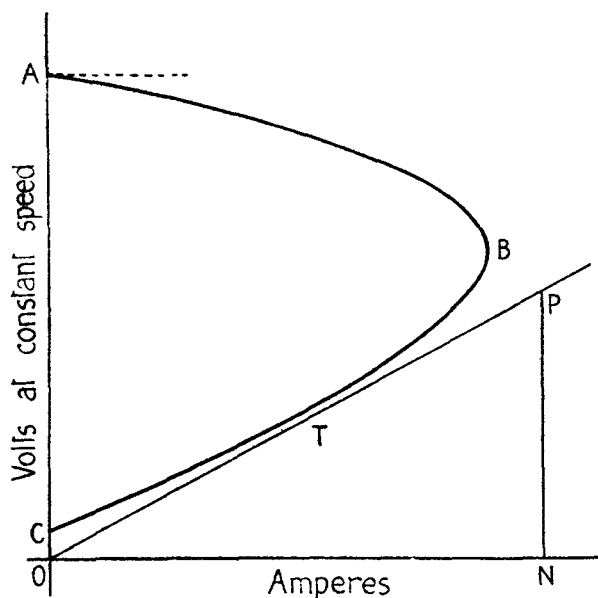


FIG. 22, CHAP. IV.—External characteristic, shunt-wound generator.

Series-wound generator

36. In this machine the field current (neglecting the small fraction taken by the field regulator if fitted) is the same as the armature current, and the E.M.F. increases as the field current increases, i.e. as the external resistance is reduced. For low output currents, the increase in terminal P.D. is proportional to the increase in armature current, but near full load this proportionality fails owing to the increasing reluctance of the magnetic circuit at high flux density. As shown by the dotted line in fig. 23, the E.M.F. tends to become constant, but the terminal P.D. reaches a maximum value and then falls, because the increase of E.M.F. due to the stronger field is less than the increased fall of P.D. in the field magnet and armature windings. The terminal P.D. therefore varies with the load current in the manner shown by the curve OBC. The corresponding values of load current and terminal P.D. for any given external resistance may be found by drawing a line such as OR, having a slope equal to the stipulated external resistance. With the external resistance corresponding to OR, the machine will have very little excitation, and in fact will be very unstable unless the resistance is considerably reduced. If however the external resistance is such that the corresponding line is OB, the terminal P.D. will be a maximum, and slight changes of output current will cause a slight fall in P.D. In practice it is usual to operate the generator on the portion BC of the curve. It is also seen that the generator may fail to excite if started up on open circuit.

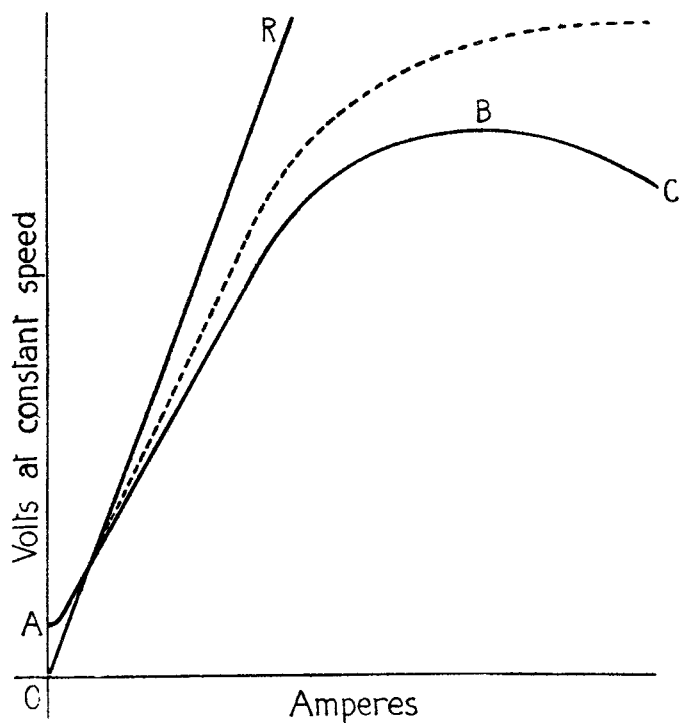


FIG. 23, CHAP. IV.—External characteristic, series-wound generator.

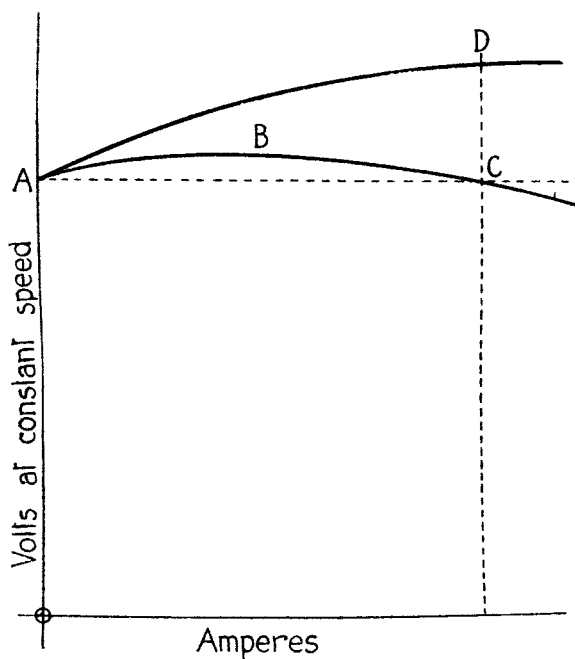


FIG. 24, CHAP. IV.—External characteristic, compound-wound generator.

The compound-wound generator

37. Since the effect of an increase of load is to reduce the voltage of a shunt-wound generator and to increase the voltage of a series-wound one (unless very heavy currents are taken by the external circuit), a combination of the two windings renders it possible to design a generator which will have the same terminal P.D. at full load as at no load. The external characteristic of such a generator is given in fig. 24, by the line ABC. At loads below full load the terminal P.D. will be slightly higher, and at overload, slightly lower than the P.D. at no load. The point C, which has been taken as full load, can be made to correspond with any desired current by suitably proportioning the series and shunt turns, and the machine is said to be level-compounded for this particular current. By increasing the number of series turns above that required to give level compounding, the generator may be designed to give a rising characteristic such as AD, and is then said to be over-compounded.

Reversal of rotation

38. If a separately excited or permanent magnet field machine is driven in the reverse direction, it generates its normal E.M.F., but its polarity will be reversed. If a shunt-wound generator is rotated in the reverse to its normal direction, however, the machine will fail to excite, for the initial E.M.F., which is set up by the residual magnetism, will be reversed, and a current will flow in the field winding which will annul the residual flux instead of increasing it. In order to cause the machine to generate when rotated reversely, the connections between field windings and brushes must be reversed, or alternatively the brush rocker must be shifted through one pole pitch. This applies also to series and compound machines.

Reversal of polarity

39. As the polarity of both series- and shunt-wound machines is dependent solely upon the polarity of the residual magnetism, no change of connections is necessary to cause a reversal of polarity. All that is necessary is to supply the field winding with current from some external source, ensuring that the residual magnetism is entirely reversed. On running the machine in the usual way the E.M.F. will then build up with its new polarity. This principle is of importance because it has a bearing upon the consequences of an accidental reversal of current in the circuit, e.g. if the external circuit contains a source of E.M.F. circumstances may arise in which the latter becomes greater than the voltage generated, and the direction of the armature current would then be reversed. In a series-wound generator, the field current will also be reversed, and hence the polarity will be changed. The external and generated E.M.F.'s are then in the same direction and excessive currents will flow, almost certainly causing damage. In a shunt-wound generator, however, the direction of the field current will not be reversed and the two E.M.F.'s remain in opposition. The armature current will therefore not be excessive and on the removal of the excessive external E.M.F., the currents will resume their original directions. The external E.M.F. alluded to may be a bank of accumulators under charge, and it will be seen that the shunt generator is suitable for this purpose. The series generator cannot be used for accumulator charging without complicated switch gear, because it will not excite on open circuit. The compound generator is unsuitable for battery charging owing to the possibility of the above effect taking place, due to the influence of the series turns.

Losses in D.C. generators

40. The various power losses which occur may be divided into—

- (i) Copper losses, due to the electrical resistance of the windings, both field and armature.
- (ii) Iron losses, due to hysteresis and eddy currents.
- (iii) Mechanical losses, due to friction between (a) shaft and bearings, (b) commutator and brushes, (c) the rotating armature and the surrounding air. The latter is often called the "windage loss".

CHAPTER IV.—PARAS. 41-42

The copper losses can be calculated if the resistance of the various windings and the current carried by them is known, e.g. the field winding copper loss is $I_f^2 R_f$, where I_f is the field current and R_f the resistance of the field winding. R_f should be measured when the winding has reached its working temperature. The iron and friction losses cannot be calculated easily or accurately, but it is possible to measure the total losses, and the iron losses may then be obtained from these by deducting the calculated copper loss. The efficiency of a generator is the ratio $\frac{\text{power output of the machine}}{\text{power supplied}}$. It is obviously less than unity but is generally expressed as a percentage, i.e. multiplied by 100.

41. It is usual to distinguish between three different efficiencies, known as

- (i) the commercial efficiency, which is the ratio $\frac{\text{power in the external circuit}}{\text{power supplied by the mechanical source}}$,
- (ii) the electrical efficiency, which is the ratio $\frac{\text{power in the external circuit}}{\text{power in external circuit} + \text{copper losses}}$,
- (iii) the mechanical efficiency, which is the ratio $\frac{\text{power in the external circuit} + \text{copper losses}}{\text{power supplied by the mechanical source}}$.

The manner in which the total input power is dissipated is shown in the following table :—

Power input = B.H.P. $\times$ 746.	
Bearing and brush friction losses.	Power supplied to armature.
Windage and iron losses.	Electrical power developed.
Armature and brush resistance losses.	Power output of armature.
Field resistance losses.	Power supplied to external circuit.

Provided that certain data are known the efficiencies may be calculated in the manner explained below.

42. The quantities which can be directly measured are (i) the H.P. P_m supplied to the generator, (ii) the terminal voltage V , (iii) the current output (I), (iv) the resistances of the armature (R_a) and field-winding (R_f). The efficiencies are expressed in terms of these known quantities.

Let the generated E.M.F. be E volts, the armature current I_a and the field current I_f .

In a series-wound generator $I_a = I_f = I$, while $E = V + I(R_a + R_f)$. The following relations then hold :—

Total watts generated = $VI + I^2(R_a + R_f)$.

Mechanical power supplied = P_m (H.P.)
 $= P_m \times 746$ watts.

Watts supplied to external circuit = VI .

$$\text{Hence the mechanical efficiency} = \frac{VI + I^2(R_A + R_F)}{P_M \times 746}.$$

$$\text{The electrical efficiency} = \frac{VI}{VI + I^2(R_A + R_F)}.$$

$$\text{The commercial efficiency} = \frac{VI}{P_M \times 746}.$$

In a shunt-wound generator $I_A = I + I_F = I + \frac{V}{R_F}$ while $E = V + I_A R_A = V + R_A \left(I + \frac{V}{R_F} \right)$ and the corresponding relations are :—

$$\text{Total watts generated} = EI_A$$

$$= \left\{ V + R_A \left(I + \frac{V}{R_F} \right) \right\} \left\{ I + \frac{V}{R_F} \right\}$$

$$= V \left(I + \frac{V}{R_F} \right) + R_A \left(I + \frac{V}{R_F} \right)^2$$

$$\text{Mechanical efficiency} = \frac{V \left(I + \frac{V}{R_F} \right) + R_A \left(I + \frac{V}{R_F} \right)^2}{P_M \times 746}$$

$$\text{Electrical efficiency} = \frac{VI}{V \left(I + \frac{V}{R_F} \right) + R_A \left(I + \frac{V}{R_F} \right)^2}$$

$$\text{Commercial efficiency} = \frac{VI}{P_M \times 746}$$

43. *Examples.*—(i) What horse-power is required to drive a 250 kilowatt dynamo when it is developing its full power, if the commercial efficiency at full load is 92 per cent. ?

$$\frac{92}{100} = \frac{250,000}{P_M \times 746} \text{ watts}$$

$$\therefore P_M = \frac{250,000}{746} \times \frac{100}{92}$$

$$= 364 \text{ H.P.}$$

(ii) A shunt-wound generator is supplying 250 amperes at 440 volts terminal P.D. Armature winding = 0.058 ohm, shunt winding = 81.7 ohms. Determine the armature current, the generated E.M.F. and the electrical efficiency.

$$\text{Field current} = \frac{440}{81.7} = 5.4 \text{ amps.}$$

Armature current = 255.4 amperes.

Total E.M.F. = 440 + 255.4 × 0.058.

$$= 440 + 14.8$$

$$= 454.8 \text{ volts}$$

$$\text{Electrical efficiency} = \frac{440 \times 250}{454.8 \times 255.4}$$

$$= 94.7 \text{ per cent.}$$

CHAPTER IV.—PARA. 44

(iii) If the friction and iron losses together are 3 per cent. of the output, find the commercial efficiency of the above generator.

$$\text{Output} = 440 \times 250 \text{ watts.}$$

$$\begin{aligned}\text{Friction and iron losses} &= \frac{3}{100} \times 440 \times 250 \\ &= 3 \times 44 \times 25 \text{ watts} \\ &= 3 \times 11 \times 100 \text{ watts} \\ &= 3,300 \text{ watts}\end{aligned}$$

$$\begin{aligned}\text{Input power} &= 454.8 \times 255.4 + 3,300 \\ &= 116,000 + 3,300 \\ &= 119,300 \text{ watts}\end{aligned}$$

$$\begin{aligned}\text{Output power} &= VI = 440 \times 250 \\ &= \frac{440,000}{4} = 110,000\end{aligned}$$

$$\begin{aligned}\text{Commercial efficiency} &= \frac{110,000}{119,300} \\ &= 92 \text{ per cent.}\end{aligned}$$

Direct current motors

44. Any dynamo-electric generator can be used as an electric motor, that is, if the machine is supplied with electrical energy it will convert a proportion of it into mechanical energy. This is the reverse of its action as a generator when it receives mechanical energy, a portion of which is converted into electrical energy. The action of a motor depends upon the force exerted between a magnetic field and a current-carrying conductor (Chapter II). It has been shown that the mutual force between the field and the conductor is equal to $\frac{BI}{10}$ dynes per centimetre, where B is the flux density of the field and I is the current in the conductor, in amperes. If the flux is fixed in space and the conductor is free to move the direction of motion of the latter is given by the left hand rule. In an actual wound armature as described above, all the conductors under one pole carry currents in the same direction, while all those under an opposite magnetic pole carry currents in the reverse direction. All the forces exerted tend to rotate the armature in the same direction, and a continuous electro-dynamic torque is maintained.

Examples.—(i) A conductor 50 centimetres in length carrying a current of 10 amperes is situated in a field in which $B = 10,000$ gauss. Find the force exerted upon it, in pounds.

$$\begin{aligned}f &= \frac{BI}{10} \text{ dynes} \\ &= \frac{10,000 \times 50 \times 10}{10} = 500,000 \text{ dynes} \\ &= \frac{500,000}{981} \text{ grams} \\ &= \frac{500,000}{981 \times 453} \text{ lb.} \\ &= 1.125 \text{ lb.}\end{aligned}$$

(ii) Find the torque in pounds-feet on an armature having 2,000 conductors, each carrying 25 amperes, the diameter of the armature being 1 metre, its length 0.5 metre, and the flux density in the airgap 5,000 gauss. Assume that the poles embrace 70 per cent. of the surface of

the armature. The number of conductors situated in the field is 70 per cent. of the total or $\frac{70 \times 2,000}{100} = 1,400$. Each of these carries $\frac{100}{4}$ amperes, and the total ampere-conductors is $1,400 \times \frac{100}{4} = 35,000$. The total force acting on all the conductors is $\frac{BLIN}{10} = 5,000 \times 50 \times \frac{35,000}{10}$ dynes.

$$= \frac{5,000 \times 50 \times 3,500}{981 \times 453} \text{ lb.}$$

$$= 1,970 \text{ lb.}$$

The torque or turning moment is the force multiplied by the radius on which it is exerted. The radius is 0.5 metre = 50 centimetres = $\frac{50}{2.54 \times 12}$ feet.

$$\text{Torque} = \frac{50}{2.54 \times 12} \times 1,970 \text{ pounds-feet}$$

$$= 3,220 \text{ pounds-feet.}$$

Back E.M.F. and armature reaction

45. When the armature is in rotation, its conductors cut the flux, and an E.M.F. is induced in them, exactly as is the case when the rotation is caused by some external source of energy. By Lenz's law this induced E.M.F. is in such a direction that it opposes the force which produces it, and as the rotation is caused by the current flowing in the armature conductors, the induced E.M.F. is in opposition to the applied E.M.F. and is generally referred to as the "back E.M.F.". In any given machine, running as a motor, the back E.M.F. has the value which the generated E.M.F. would have if the machine were driven mechanically, i.e. as given by the formulae of paragraph 25, hence the back E.M.F. varies directly as the speed of the machine. Now we have seen that in the motor, the electro-dynamic torque causes rotation, whereas the effect of the back E.M.F. is to reduce the armature current and therefore to decrease the torque. Friction and air resistance also oppose rotation, and the torque available to cause acceleration of the armature decreases as the speed increases. At some particular speed the driving torque is exactly equal to the torque opposing motion, and the armature receives no further acceleration, but continues to run at this steady speed. As in the generator, the flux caused by the armature currents combines with the stator flux to produce a resultant field which is not symmetrical about the geometrical neutral axis, but as the current is flowing in the opposite direction to the generated (i.e. back) E.M.F. the magnetizing effect of the armature windings is opposite to that in a generator, and the total magnetic field is distorted against the direction of rotation. The brushes must be given a lag or trail to obtain satisfactory commutation, unless inter-poles are used for this purpose. It is seen on reflection that if inter-poles are used to obtain sparkless commutation, any inter-pole must have the polarity of the main pole which is behind it in the direction of rotation.

Motor characteristics

46. Any form of field excitation may be used for a motor, and the behaviour of the various forms of motor may be compared by means of their mechanical characteristics, the latter being curves showing the relation at constant supply voltage between the speed and the torque. It is desirable to bear in mind that the back E.M.F. is normally very nearly equal to the applied E.M.F. and so is nearly constant. As $E_b = K n \Phi$ the product $n \Phi$ is very nearly constant, or the speed n varies (approximately) inversely as the flux.

Shunt-wound motor

47. The mechanical characteristic of this type is shown in fig. 25. The field current is constant because the terminal P.D. is constant. It is seen that as the torque is increased the

CHAPTER IV — PARA. 48

speed falls, because the increase of torque is due to an increased demand for mechanical power, i.e. an increase of mechanical load, and therefore an increase of armature current, which can only be obtained by a decrease of back E.M.F., and consequently a lower speed. Shunt-wound motors are suitable for driving machine tools and motor-generators.

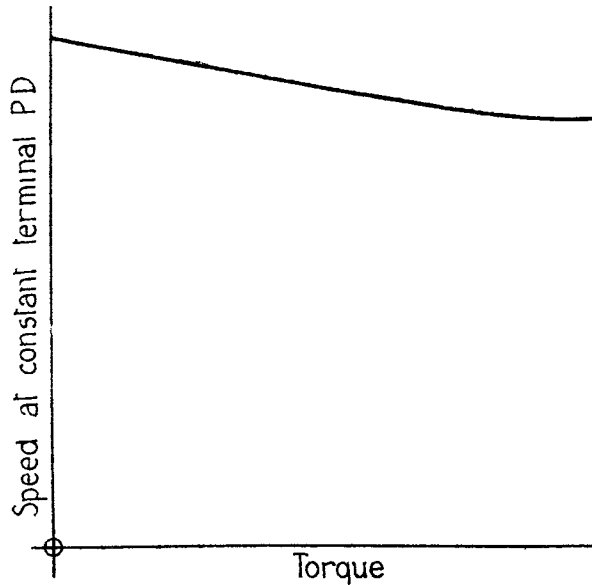


FIG. 25, CHAP. IV.—Mechanical characteristic of shunt-wound motor

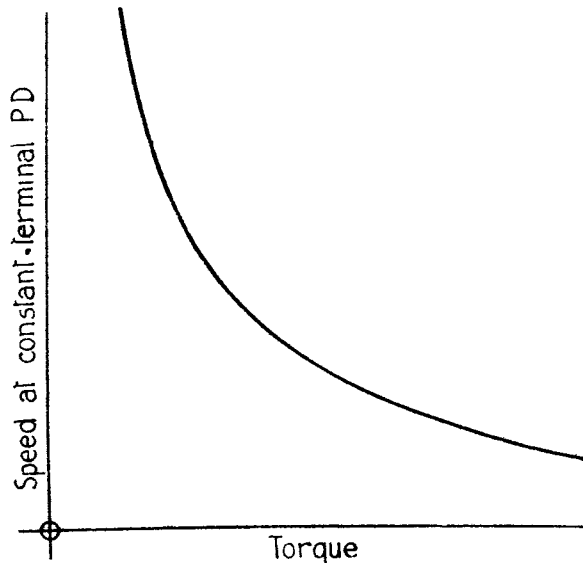


FIG. 26, CHAP. IV.—Mechanical characteristic of series-wound motor.

Series-wound motor

48. The mechanical characteristic is shown in fig. 26. In this machine an increase of torque calls for an increase of armature current as usual, but as this current also flows through the field

winding, the field is strengthened and the speed decreases. Series motors are suitable for traction purposes, because they exert their greatest torque at low speeds, i.e. when starting. As the load decreases they increase speed automatically. They are also adapted for hoisting purposes, for while a shunt-wound motor would raise any load whatever at practically the same speed, a series motor will raise a light load at high speed, but a heavy load at low speed. Series motors should not be used in circumstances where the load can be totally removed, for there is then a danger that the speed may become excessive and cause damage to the armature by the large centrifugal forces then developed. Fans are generally driven by series motors.

Compound-wound motors

49. A motor having both series and shunt field windings may have them connected in either of two ways. If their magnetizing effects are in the same direction, i.e. if the series and shunt currents flow in the same direction round the pole, the machine is said to be cumulatively wound while if the series and shunt currents are in opposition the machine is said to be differentially wound. In the latter type the series turns may be made to weaken the field by an amount just sufficient to make the speed at some particular load the same as at no load. The speed at other loads will then be slightly different, the mechanical characteristic being somewhat as shown in fig. 27a. This type of motor is rarely used. When a compound-wound motor is referred to without qualification the cumulative type is implied. The mechanical characteristic, fig. 27b, is intermediate between those of the shunt and series motors. This type of machine is used where a series characteristic is desired, e.g. for hoisting purposes, but where the load may be entirely removed on occasion. The effect of the shunt winding is to prevent a dangerous increase of speed when this occurs.

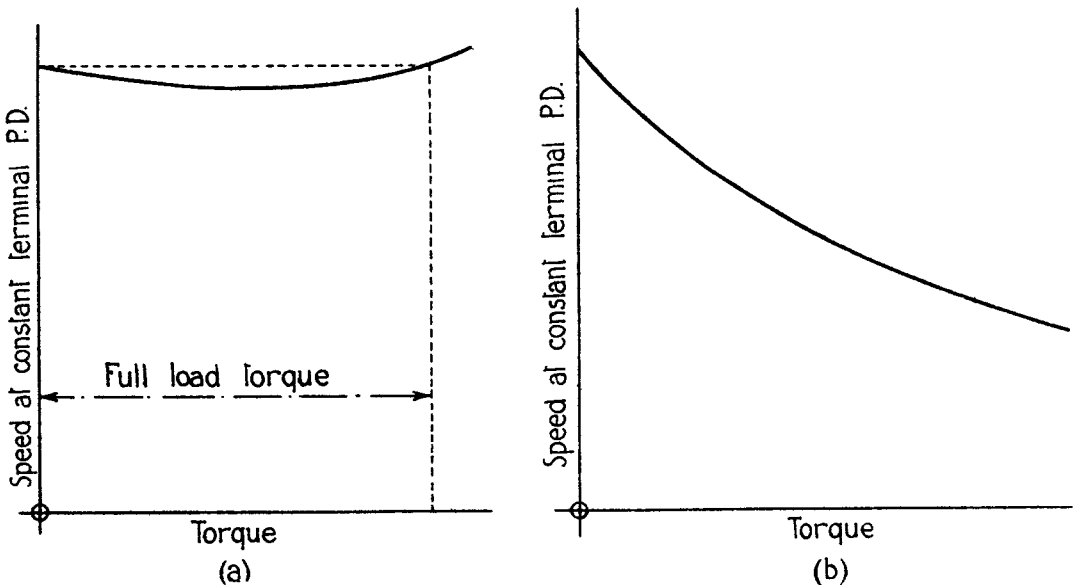


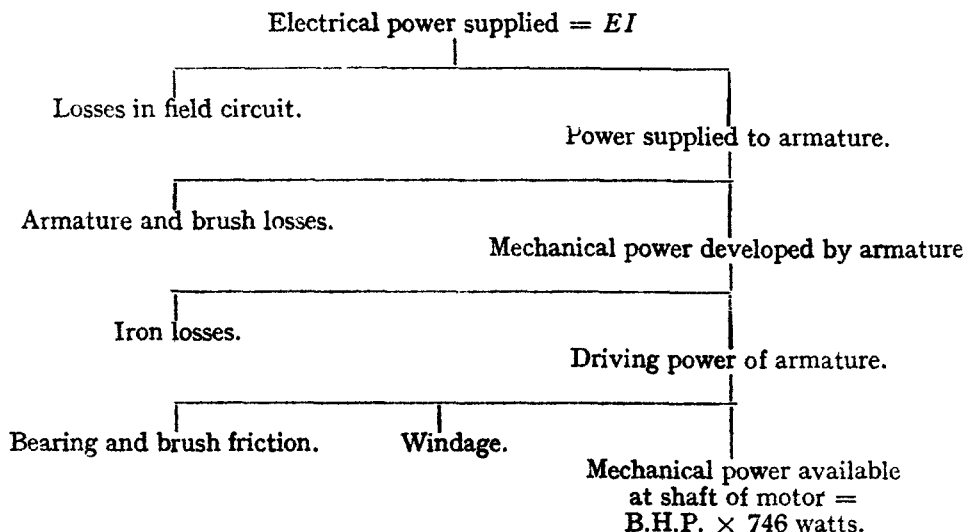
FIG. 27, CHAP. IV.—Mechanical characteristics, compound-wound motors.

Losses in motors

50. The various losses which occur in electric motors are identical with those which occur in generators, i.e. copper losses, iron losses and frictional losses. The commercial efficiency is the ratio $\frac{\text{power obtained from the motor}}{\text{electrical power supplied}}$, both numerator and denominator being expressed in watts.

CHAPTER IV.—PARAS. 51–52

The electrical efficiency is the ratio $\frac{\text{mechanical power developed by armature}}{\text{electrical power supplied}}$, and the mechanical efficiency is the ratio $\frac{\text{power obtained from motor}}{\text{mechanical power developed in armature}}$. As before the power balance sheet may be shown in tabular form thus :—



Motor starters

51. When the armature of a motor is at rest it obviously develops no counter-E.M.F. and if it were switched directly on to the mains, an extremely large current would flow, causing serious damage to the machine. Thus, a 10 H.P. series-wound 220-volt motor may have a normal full load current of about 40 amperes, an armature resistance of 0·2 ohm, and a field resistance of 0·1 ohm. If directly connected to the mains, the initial current would be $\frac{220}{0·3} = 733$ amperes which is practically twenty times the normal full load current, and would almost certainly do serious damage to the armature. To avoid this a series starting resistance is used, this resistance being cut out step by step as the speed of the motor, and the counter-E.M.F., increases. Small series motors usually have a field winding of sufficiently high resistance and inductance to render the use of a starting resistance optional.

Shunt motor starter

52. In all except the smallest sizes a starting resistance is essential for a shunt motor, because the field winding is in parallel with the armature, hence the latter would be thrown straight on the supply mains if a simple make-and-break switch were employed; owing to the extremely low resistance, a very heavy current would then flow through the windings and would cause severe damage. This starter also generally embodies two safety devices known as the overload release and no-volt release respectively. The general features of a shunt motor starter designed for manual operation is shown in fig. 28. The moving arm (A) which is normally held in the off position by a spring (S), is arranged to sweep over a row of contacts which are arranged on an arc. To these contacts are connected a series of tappings on the resistance R, which is in series with the armature only, and on moving the arm to the first contact (1) the full mains voltage is applied to the field winding, causing the field to build up to its full excitation, but allowing only a limited current to flow through the armature. This may actually be as much as double the normal running current. The armature then starts to rotate and generates a back E.M.F., the armature

current falling as the speed increases. When the armature current has fallen to its normal running value, the arm is moved to the second contact (2). The armature current again rises, but as the speed increases the additional back E.M.F. soon reduces the current to normal once more. The operation is repeated in this manner until the whole of the resistance is cut out and the armature is then connected straight across the supply mains. When the arm is in its final position it is held in place by an electro-magnet acting upon a "keeper" which is attached to the moving arm, and the spring (S) is arranged to pull the arm back into the "off" position should the current through the field winding fail. The winding of this electro-magnet is connected in series with the field winding. This electro-magnet is called the no-voltage release; its inclusion ensures (i) that the switch arm will return to its off position should the supply voltage fall below a certain value, and (ii) that the motor will be switched off automatically in the event of a break occurring in the field circuit of the machine. The latter precaution is necessary because otherwise, the sudden weakening of the field would cause the motor to increase in speed to a dangerous extent. The overload release or overload cut-out is an electro-magnet the winding of which is directly in series with the supply. A piece of soft iron which is pivoted at one end is so arranged that when attracted by this electro-magnet, it short-circuits the no-voltage release coil, and the starter arm is allowed to fall back to the "off" position. The armature of the overload coil is normally held off its contacts by a small spring, the tension of which is just overcome by a given overload.

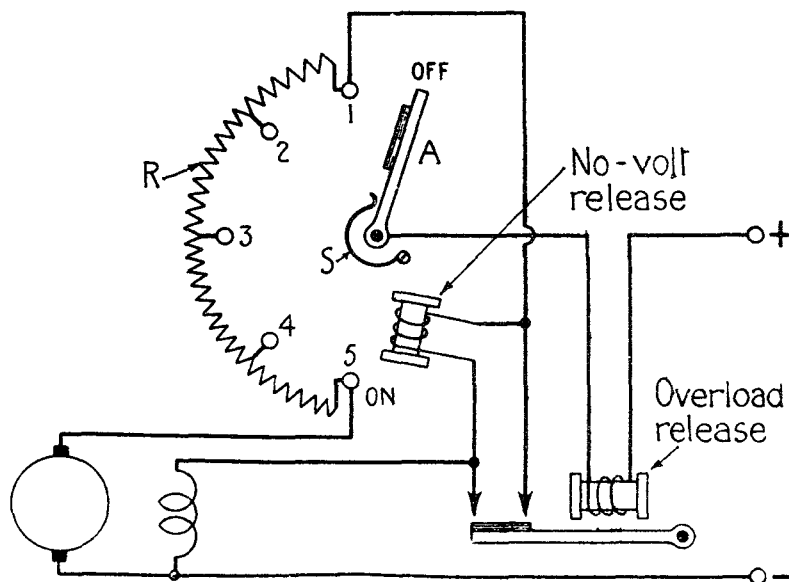


FIG 28, CHAP. IV.—Shunt motor starter.

Voltage transformation

53. It frequently happens that the voltage required for a certain purpose differs from that which is immediately available. If the latter is higher than the required voltage, a simple expedient is to insert an appropriate resistance in series with the low voltage appliance, so that the correct voltage is applied to the terminals of the latter. The wastefulness of this procedure is not always appreciated, and an example may be of assistance to the reader. Consider the charging of a large capacity secondary battery from 220-volt mains. The nominal voltage of the battery being 20 volts, i.e. 10 cells, we will assume that it has fallen to 1.8 volts per cell or 18 volts, and the internal resistance of the battery to be 0.1 ohm, while its charging rate is 10 amperes. The necessary value of series resistance to give the desired charging current is found by Ohm's law,

CHAPTER IV.—PARA. 53

$$I = \frac{E - E_b}{R}$$

where I is the charging current, E the supply voltage, E_b the E.M.F. of the battery and R the required resistance. Transposing and inserting numerical values

$$R = \frac{220 - 18}{10} \\ = 20.2 \text{ ohms.}$$

Now consider the electrical efficiency of the charging process. The power supplied is $E I$ watts, the power converted into heat is $I^2 R$ watts and the efficiency is $\frac{E I - I^2 R}{E I}$ or $\frac{E - IR}{E}$ which is

$$\frac{220 - 202}{220} = \frac{18}{220} \text{ or } 8.16 \text{ per cent.}$$

Of the cost of charging the battery during the whole period of service, elevenpence in every shilling is wasted, and the wastage can only be avoided by adopting some method of changing the voltage, from that of the supply to that actually required. When the supply consists of direct current and voltage, this change of voltage can be achieved by the use of either a motor generator or by some form of booster. Although the use of a series resistance enables a reduction of voltage to be obtained at the terminals of any apparatus, the use of some form of rotating machinery is essential when an increase of voltage is desired. In this respect alternating supply enjoys a considerable advantage over direct supply, for by means of a suitable A.C. transformer, which has no rotating parts and requires a minimum of attention, the voltage can be raised or lowered as may be most desirable.

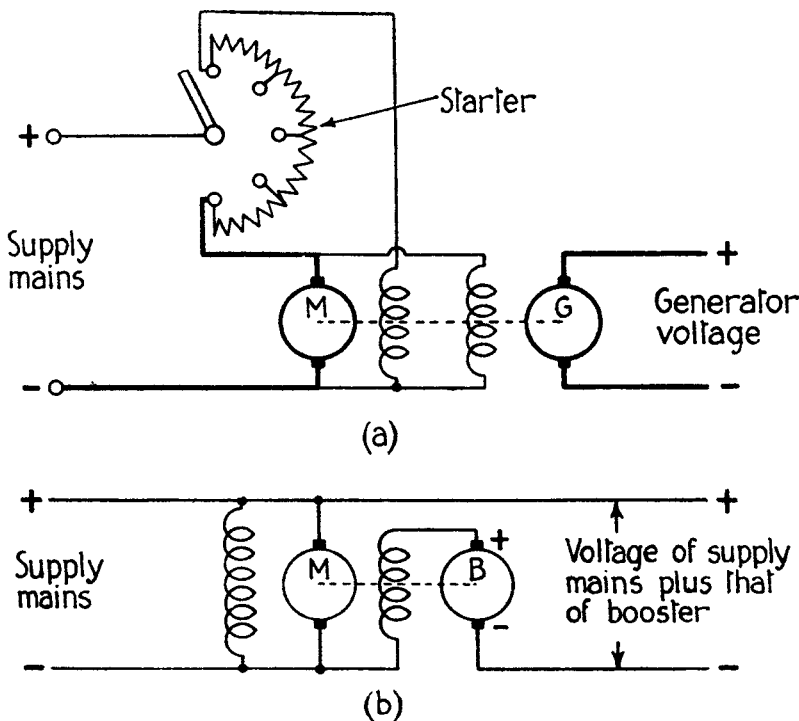


FIG. 29.—Motor generator and booster.

The motor generator

54. This consists of a motor running off the supply mains, which drives an electrical generator providing power at the desired voltage. In its simplest form the machine comprises separate motor and generator the shafts of which are coupled together, the two machines being mounted on a common bed plate. The motor is invariably shunt-wound, and the field excitation for the generator is provided from the supply. The connections to the input and output terminals are therefore as shown in fig. 29a. This type of motor generator has the advantage that the generated voltage can be controlled by variation of the generator field excitation independently of the speed of rotation, which is controlled by the motor field. Double-wound motor generators are those which carry two separate windings on the same armature core, each winding having its own commutator. It is convenient to mount one of these at each end of the armature. One winding is connected through its commutator and brush gear with the supply mains, and drives the armature as a motor, while the other winding is designed to give the desired voltage. Field excitation is supplied from the mains. The double wound machine is lighter and smaller than the separate unit motor generator, and its cost is rather lower, while the effects of armature reaction are not so pronounced, owing to the opposing reactions of the two windings. On the other hand, the generator voltage cannot be controlled by variation of the field excitation, because if the field is weakened with a view to reducing the voltage, the machine considered as a motor will increase speed, and the voltage will increase also. This disadvantage prevents the extensive adoption of the double-wound machine.

55. A booster is a generator, the voltage of which can be added to that of another generator, thus increasing the total voltage. The machine is primarily merely a motor generator, but the output voltage of the generator is connected in series with the supply voltage, the simplest arrangement being given diagrammatically in fig. 29b, from which it is apparent that the terminal voltage of the booster is the sum of that generated by the machine and that of the supply. If a reduction of voltage is required, the booster field is connected in such a manner that the E.M.F. of the booster armature is in opposition to that of the mains. These machines have little service application.

CHAPTER V.—SINGLE-PHASE ALTERNATING CURRENTS

INSTANTANEOUS, PEAK AND R.M.S. VALUES

Heating effect of sinusoidal current

1. When an alternating E.M.F. is applied to an electric circuit the immediate result is the production of an alternating current of the same frequency, and in this chapter it is proposed to consider the behaviour of such circuits. The simplest wave-form of an alternating quantity is the sine wave and unless otherwise stated it is always assumed that the current has this wave-form. The value of the current at any instant is called its instantaneous value, while the maximum value attained at any instant during each half-cycle is termed the peak value. The instantaneous value, denoted by i , is related to the peak value $\mathcal{I}$ by the equation

$$i = \mathcal{I} \sin \omega t$$

where $\omega = 2 \pi f$

The instantaneous value of an alternating current or E.M.F. is only measurable by expensive and complicated apparatus, and even if known a single instantaneous value gives no indication of the total effect of the current over a considerable period; the peak value is merely one particular instantaneous value and can only be measured in the same way. Any kind of measuring instrument which indicates the average value of the current, e.g. the moving coil ammeter, must of necessity give a zero reading if connected in an A.C. circuit, for it is obvious from inspection of the wave-form that the average value of the current over any number of complete cycles is zero. The heating effect of an electric current, however, is entirely independent of the direction in which the current is flowing, and the effective value of an alternating current is said to be I amperes if it has the same heating effect as an unvarying current of I amperes. In other words, an alternating current has the same effective value as a given direct current if both produce equal deflection of the pointer in a hot-wire ammeter, so that the latter instrument may be calibrated by comparison with a sub-standard moving coil ammeter in a D.C. circuit, and afterwards used in an A.C. circuit for current measurement. In order to find the value of direct current which is equivalent to a given alternating current $i = \mathcal{I} \sin \omega t$ therefore, it is necessary to find the heating effect of the latter. Now at any instant t , counted from the beginning of a cycle, the power expended, p , (i.e. the rate at which energy is being expended at that particular time) is $i^2 R$ joules per second, if R is the total resistance of the circuit and i is the instantaneous value of the current. Hence

$$p = i^2 R = R \mathcal{I}^2 \sin^2 \omega t$$

The current $i = \mathcal{I} \sin \omega t$, and the instantaneous values of i^2 , are plotted (side by side) over a complete period in fig. 1, in which the peak value of the current, $\mathcal{I}$, has been arbitrarily assigned the value 4 amperes. It will be observed that although i passes through both positive and negative values, the curve showing the square of the current has positive values only, which is a

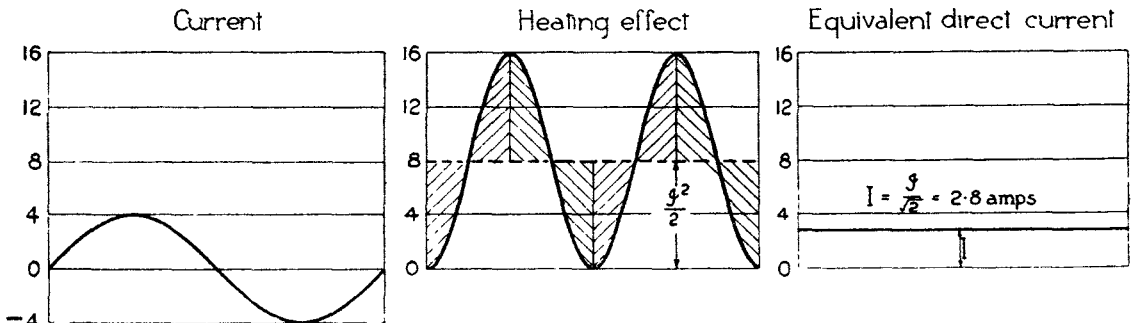


FIG. 1, CHAP. V.—Heating effect of an alternating current.

CHAPTER V.—PARA. 2

graphical illustration of the fact that the square of a real quantity is always positive. This curve contains two complete cycles in the time taken for one cycle of i , and its average height is 8 units. This may be verified by tracing the curve on to thin paper, and cutting off the shaded portions above the dotted line. The peaks thus removed will then be found to fit into the hollows left below the line, showing that the area of the whole i^2 curve is $8 \times T$ units, that is $\frac{16}{2} T$ or $\frac{\mathcal{I}^2}{2} T$. The average height of the curve is the area divided by the length of the base,

the latter being T units. Hence the average height is $\frac{\mathcal{I}^2}{2}$ units. Another way of arriving at the same conclusion is as follows. The "current-squared" curve is a cosine curve of twice the frequency of the current curve, but its axis is displaced upwards by 8 units, so that it never becomes negative in sign. The curve can therefore be represented by the equation

$$i^2 = \frac{\mathcal{I}^2}{2} - \frac{\mathcal{I}^2}{2} \cos 2 \omega t$$

the factor $2 \omega t$ signifying that the frequency of the i^2 curve is twice that of the i curve. $\frac{\mathcal{I}^2}{2}$ is the average value of i^2 , the average value of $\frac{\mathcal{I}^2}{2} \cos 2 \omega t$ over the whole curve being zero.

This may be a convenient opportunity to point out that the average value of any function such as $\sin \omega t$, $\sin \theta$, $\cos n \omega t$, etc., for any number of complete periods is always zero.

The heating effect of the current i , in the resistance of R ohms, is therefore given by the equation $P = \frac{\mathcal{I}^2}{2} R$ joules per second, which may be written

$$P = \left(\frac{\mathcal{I}}{\sqrt{2}} \right)^2 R \text{ watts.}$$

Let us now compare two currents having equal heating effect, a direct current of unvarying value I , and an alternating current of peak value $\mathcal{I}$, which are assumed to flow through a resistance of R ohms. The steady current has a heating effect of $I^2 R$ watts, and the alternating current a heating effect of $\left(\frac{\mathcal{I}}{\sqrt{2}} \right)^2 R$ watts. The heating effect of the two currents will be equal if

$$I^2 = \left(\frac{\mathcal{I}}{\sqrt{2}} \right)^2, \text{ or } I = \frac{\mathcal{I}}{\sqrt{2}}$$

The expression $\frac{\mathcal{I}}{\sqrt{2}} = \frac{\mathcal{I}}{1.414}$ or $.707 \mathcal{I}$ is called the root-mean-square or R.M.S. value of the current whose peak value is $\mathcal{I}$. Whenever an alternating current is said to have a value I amperes, without further qualification, the R.M.S. value is implied. Also, since the heating effect of a direct current is equal to $\frac{V^2}{R}$ joules per second, the R.M.S. value of an alternating voltage of

peak value $\mathcal{V}$ is $\frac{\mathcal{V}}{\sqrt{2}}$. When used in this way the factor $\sqrt{2}$ is called the peak factor. It will be observed that three conventions of notation have been introduced in the preceding discussion, the instantaneous value of an alternating quantity being denoted by a small italic letter, the peak value by a cursive or script capital, and the R.M.S. value by an italic capital. This notation will be followed throughout this chapter.

Wave-Form

2. Up to the present we have considered only those alternating quantities which obey the simple sine law, but the shape of the graph obtained by plotting the instantaneous value or displacement of the quantity at various intervals of time may take any one of an infinite variety of forms, and this displacement-time curve may be referred to as the wave-form of the quantity.

Provided that the wave-form repeats itself at regular intervals, it can be proved that it is built up by the addition of a number of simple sine waves of various frequencies. The frequency of each component sine wave is an integral multiple of some fundamental frequency (which may or may not be present), and is said to be in harmonic relation with the latter, while the higher frequencies are called the harmonics of the fundamental. If the average value of the quantity over any number of complete periods is not zero, a constant displacement must also exist; for instance, if the quantity under consideration is an alternating current, the wave-form being complex and repetitive, it may be represented by an equation of the form

$$i = I_0 + \mathcal{I}_1 \sin (\omega t + \varphi_1) + \mathcal{I}_2 \sin (2 \omega t + \varphi_2) + \mathcal{I}_3 \sin (3 \omega t + \varphi_3) + \dots$$

I_0 is the average value of the current over any number of complete periods and may be regarded as a direct current superimposed upon the alternating components. The fundamental frequency is $\frac{\omega}{2\pi}$ and this is also referred to as the first harmonic. The second harmonic has a frequency of twice the fundamental or $\frac{2\omega}{2\pi}$, while the third harmonic frequency is $\frac{3\omega}{2\pi}$ and so on. The angles represented by the symbols $\varphi_1, \varphi_2, \varphi_3, \dots$ are inserted to signify that it is not necessary for all the components to pass through zero displacement at the instant arbitrarily assumed to be zero time.

3. Frequencies which are even multiples of the fundamental frequency f , e.g. $2f, 4f$, etc. are called the even harmonics, and those which are odd multiples, $3f, 5f$, etc. are called the odd harmonics. An E.M.F. generated by rotating machinery is free from even harmonics, but may

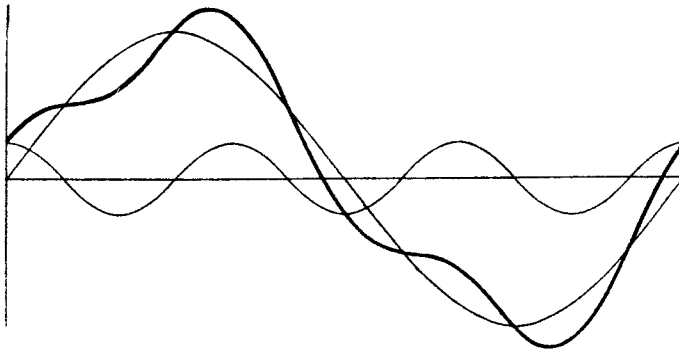


FIG. 2, CHAP. V.—Wave with third harmonic.

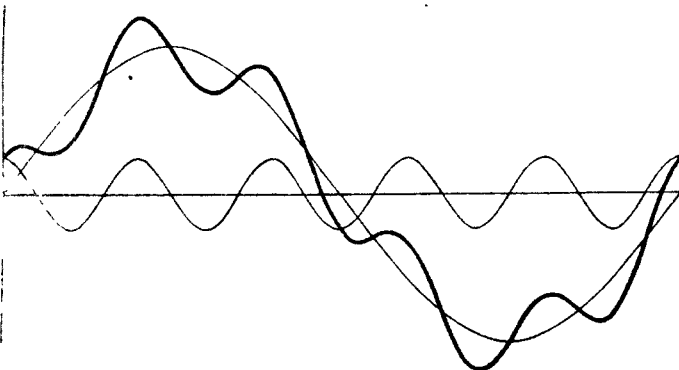


FIG. 3, CHAP. V.—Wave with fifth harmonic.

CHAPTER V.—PARA. 4

contain odd harmonics. This entails that the positive and negative half-cycles have exactly the same wave-form, which must be true if the E.M.F. is produced by rotation of a conductor in an unvarying magnetic field. The odd harmonics are produced by uneven flux distribution, and do not occur if the field is uniform as it was assumed to be when considering the production of an alternating E.M.F. in the preceding chapter. Typical wave-forms containing odd harmonics only are portrayed in figs. 2 and 3. Even harmonics are frequently found in radio circuits, a particular instance being the alternating current flowing in the anode circuit of a valve transmitter.

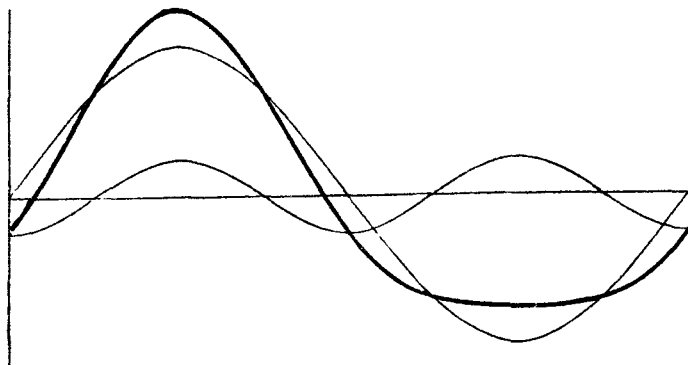


FIG. 4, CHAP. V.—Wave with second harmonic.

The existence of an even harmonic in a given wave-form may be instantly detected because its presence causes the negative half-cycles to be of a shape differing from the positive half cycles. The point is illustrated in fig. 4 which shows the result of adding to a sine wave a second harmonic which bears a certain phase relationship to the fundamental. This wave-form, with the addition of further harmonics of small peak value, is often met with. The wave-form resulting from the addition of the two components mentioned may be represented by the equation

$$i = \mathcal{I} \sin \omega t - \frac{\mathcal{I}}{2} \sin \left(2\omega t + \frac{\pi}{2} \right)$$

Effective value of current of complex wave-form

4. In order to find the heating effect of a current of non-sinusoidal but recurring wave-form we may utilise the result already obtained for the simple sine wave. If the current flows through a resistance of R ohms, the component current of fundamental frequency will cause energy to be expended at a rate of $\frac{\mathcal{I}_1^2 R}{2}$ joules per second, while the second harmonic component will cause an expenditure of energy at a rate of $\frac{\mathcal{I}_2^2 R}{2}$ joules per second and so on. The total energy expended will therefore be given by the expression

$$\begin{aligned} P &= I_0^2 R + \frac{\mathcal{I}_1^2}{2} R + \frac{\mathcal{I}_2^2}{2} R + \frac{\mathcal{I}_3^2}{2} R \dots\dots \\ &= R \left\{ I_0^2 + \left(\frac{\mathcal{I}_1}{\sqrt{2}} \right)^2 + \left(\frac{\mathcal{I}_2}{\sqrt{2}} \right)^2 + \left(\frac{\mathcal{I}_3}{\sqrt{2}} \right)^2 \dots\dots \right\} \end{aligned}$$

But this is equal to $I^2 R$, if I is the effective value of the current. Hence

$$I = \sqrt{\left\{ I_0^2 + \left(\frac{\mathcal{I}_1}{\sqrt{2}} \right)^2 + \left(\frac{\mathcal{I}_2}{\sqrt{2}} \right)^2 + \left(\frac{\mathcal{I}_3}{\sqrt{2}} \right)^2 \dots\dots \right\}}$$

or if I_1, I_2, I_3 etc. are the R.M.S. values of the currents of peak value $\mathcal{I}_1, \mathcal{I}_2, \mathcal{I}_3$ etc.

$$I = \sqrt{\left\{ I_0^2 + I_1^2 + I_2^2 + I_3^2 \dots\dots \right\}}$$

The peak factor for the complex wave is defined in the same way as before, being the ratio of the peak value to the R.M.S. value. The mean value of a single half-cycle of an alternating current is sometimes required. Consider first the simple sine wave; we may obtain an approach to the average value of $\sin \theta$, from $\theta = 0$ to $\theta = 90^\circ$, by means of a table of sines, adding up the whole series and dividing by the number of values given in the table. This is a laborious process, but if actually performed it will be found that the average value of $\sin \theta$ from 0 to 180° is very near to $\cdot 637$. Actually the calculation can be performed with much less labour, although the mathematical ideas involved are more complicated. The result of such a calculation gives the average value of $\sin \theta$ during one half-cycle as $\frac{2}{\pi}$ which is $\cdot 6366$. Hence the average value of a sine wave over one half-cycle is $\frac{2}{\pi}$ times the peak value. The ratio $\frac{\text{R.M.S. value}}{\text{mean value}}$ is called the form factor, and is equal to $1\cdot 11$ for a sinusoidal wave.

For any wave-form whatever, the average value per half-cycle may be obtained by drawing the wave to some convenient scale, calculating its area using the mid-ordinate method or Simpson's rule, and dividing the area by the length of the base line which represents one half period. An approximation to the average and R.M.S. values may also be obtained from the instantaneous values at a number of equal intervals during the cycle, as in example (2) below.

Examples.—(1) An alternating current is represented by the equation

$$i = 250 \sin \omega t + 125 \sin 3 \omega t + 50 \sin 5 \omega t.$$

Find its R.M.S. value.

Here $I_0 = 0$, $I_1 = 250$, $I_2 = 0$, $I_3 = 125$, $I_4 = 0$, $I_5 = 50$.

$$\begin{aligned} I &= \sqrt{\frac{250^2}{2} + \frac{125^2}{2} + \frac{50^2}{2}} \\ &= \sqrt{31250 + 7812\cdot 5 + 1250} \\ &= \sqrt{40312\cdot 5} \\ &= 201 \text{ amperes (nearly).} \end{aligned}$$

(2) An alternating voltage is found to pass through the instantaneous values given in the following table.

Angle	..	0°	15°	30°	45°	60°	75°	90°
Voltage	..	0	25	50	88	130	175	225
Angle	..	105°	120°	135°	150°	165°	180°	195° etc.
Voltage	..	270	230	165	80	30	0	-25 etc.

the negative half-cycle being of the same shape as the positive.

Find an approximation to the R.M.S. and average values, and the peak and form factors.

The average value of one half-cycle is

$$\begin{aligned} \frac{1}{12} [0 + 25 + 50 + 88 + 130 + 175 + 225 + 270 + 230 + 165 + 80 + 30] \\ = \frac{1468}{12} = 122\frac{1}{3} \text{ volts.} \end{aligned}$$

The ratio $\frac{\text{mean value}}{\text{peak value}}$ is $\frac{122\frac{1}{3}}{270} = \cdot 453$ which may be compared with the value $\cdot 637$ for a sinusoidal wave. The R.M.S. value is the square root of the mean of all the squares of the above ordinates, i.e.

$$\begin{aligned} V^2 &= \frac{1}{12} [0 + 625 + 2500 + 7744 + 16900 + 30625 + 50625 + 72900 + 52900 \\ &\quad + 27225 + 6400 + 900] \\ &= \frac{270344}{12} = 22529 \\ V &= \sqrt{22529} \\ &= 150\cdot 9 \text{ volts.} \end{aligned}$$

CHAPTER V.—PARAS. 5-7

The peak factor, $\frac{\text{peak value}}{\text{R.M.S. value}} = \frac{270}{150.9} = 1.79$ which again should be compared with the value 1.414 for a sinusoidal wave.

The form factor, $\frac{\text{R.M.S. value}}{\text{mean value}} = \frac{150.9}{122.3} = 1.23$. If the wave-form is plotted from the data it will be found to be more peaked than a sine wave. Such a wave has high values for its peak and form factors while the converse is true of flat-topped waves.

MEASUREMENT OF CURRENT, VOLTAGE AND FREQUENCY

Ammeters and voltmeters

5. From the foregoing, it will be appreciated that any type of ammeter or voltmeter in which the deflection is proportional to the average value of the current or voltage is unsuitable for use in A.C. circuits. Hot-wire ammeters, thermo-ammeters and moving iron ammeters may be used for measurement of alternating current, and hot-wire and electrostatic voltmeters for the measurement of alternating P.D. In general it may be stated that instruments which depend for their action upon any form of permanent magnet are unsuitable for use in A.C. circuits. Instruments depending upon magnetisation by the current, for example moving iron instruments, can be used for A.C. measurement, but special care is necessary in the design, in order to reduce the effects of eddy currents. This necessitates the sub-division of any metallic parts in proximity with the current-carrying conductors, and may be explained with reference to the repulsion type of ammeter. If the coil is wound upon a metal former, it is necessary that the former should be cut completely in a radial direction, so that eddy currents cannot circulate completely round the coil former. The iron portions may also be laminated for the same reason. Theoretically, if such an instrument is calibrated with steady current its scale should read R.M.S. value when connected in an A.C. circuit, but in practice this is not quite true, partly on account of the varying permeability of the iron with varying magnetising current, and partly owing to the inductance of the instrument which has no effect when the instrument is used to measure direct current. It is therefore desirable that such instruments should be calibrated by alternating current of the wave-form and frequency with which they are to be employed.

6. Hot-wire instruments have the great advantage that the deflection is independent of wave-form and frequency, provided that the resistance of the instrument is the same at all frequencies. It may be assumed that a hot-wire instrument calibrated at say 250 cycles per second, may be used without fear of serious inaccuracy on frequencies between 25 and 500 cycles per second, but may be in serious error at higher frequencies.

Dynamometer instruments are moving coil instruments of special design. Instead of being established by a permanent magnet as in the instruments commonly termed "moving coil type" the magnetic field is supplied by a fixed winding carrying an alternating current, this winding being connected in series with that of the moving coil. These instruments are expensive, and offer no advantages over hot-wire instruments as ammeters or voltmeters, but by a simple modification the principle is employed in the construction of one form of watt-meter, which is described later.

Frequency meters

7. Two forms of frequency meter are found in low (or commercial) frequency circuits, namely the tuned reed pattern and the induction type. In the tuned reed pattern (fig. 5) a number of steel strips are so adjusted that each vibrates at a particular frequency. A laminated soft iron core carries a magnetising winding consisting of a great many turns of fine insulated wire, the winding being connected across the supply mains in the same way as a voltmeter, and the steel strips or reeds are arranged in such a manner that they are acted upon by this electromagnet. Any reed which is adjusted to the frequency of the supply at any particular

instant is caused to vibrate with appreciable amplitude, although reeds of different frequency are scarcely affected. To facilitate observation, each reed carries at its free end a small rectangular metal flag, and the frequency is obtained by noting on the adjacent scale the frequency corresponding to the reed having maximum amplitude of vibration. The induction type (fig. 6)

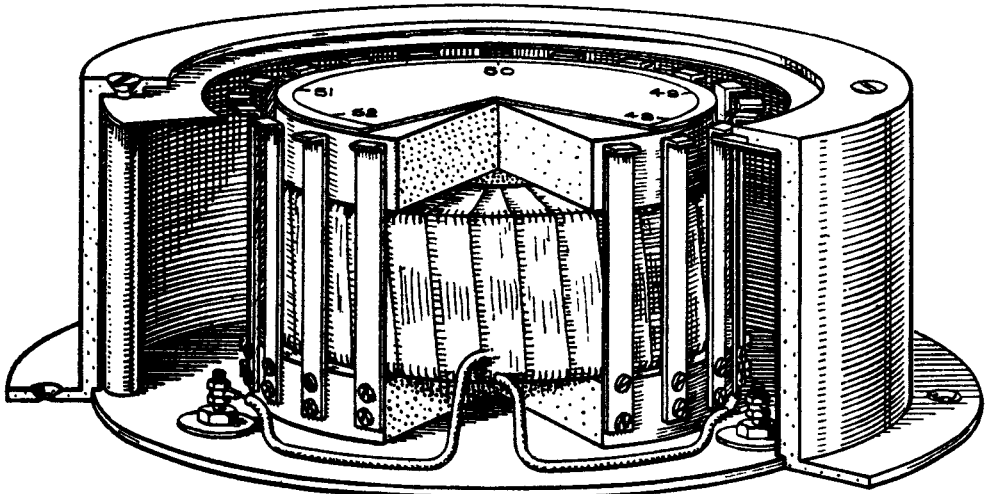


FIG. 5, CHAP. V.—Frequency meter, tuned reed pattern

consists of two coils which are mounted at right-angles and on a common axis. One of the windings has connected in series a large non-inductive resistance, while the other has a low resistance coil of large inductance in series. The two elements thus formed are each connected across the supply mains. These two coils co-operate in the establishment of a magnetic field in the space enclosed by the coils, and the direction of this field depends upon the ratio of

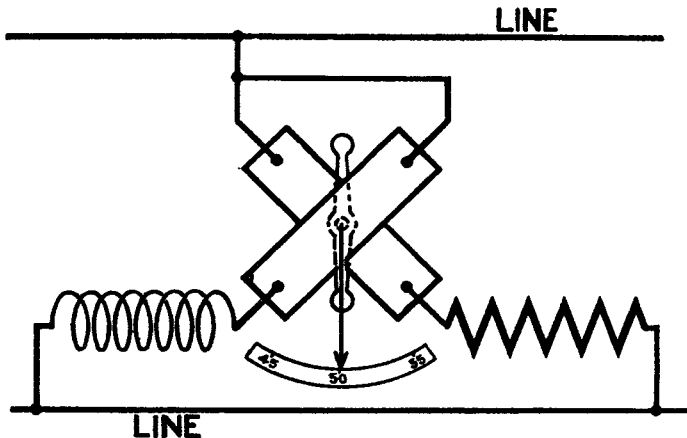


FIG. 6, CHAP. V.—Frequency meter, induction pattern.

current in the two windings. As will be seen after a study of the following sections, a change of frequency will make little or no difference to the current in the predominantly resistive branch, but the current in the winding which is predominately inductive will vary inversely as the frequency. Hence the direction of the resulting field varies with frequency, and this direction

CHAPTER V.—PARA. 8

will be taken up by a soft iron needle which is mounted upon a spindle on the common axis of the coils. A pointer attached to the spindle moves over a scale which is graduated in cycles per second.

The term frequency meter is also sometimes employed to denote an instrument which is used to measure very high frequencies (of the order of 10^4 cycles and above). Such instruments are described in a later chapter.

REPRESENTATION OF ALTERNATING QUANTITIES BY VECTORS

Vector quantities

8. A vector quantity is one which has direction as well as magnitude, for example, the magnetic field strength H at any point in the field is a vector quantity. Those quantities which have magnitude only, e.g. work, are called scalar quantities. A vector quantity may be represented in magnitude and direction by a straight line, and such a line is often referred to as "the vector" representing the quantity. When used in connection with alternating quantities, vectors are used in a somewhat special manner. The line is supposed to be fixed at one end and to rotate at the frequency of the alternation. For example, consider a voltage $v = \mathcal{V} \sin \omega t$. This voltage may be represented by a line of length $\mathcal{V}$ units, rotating in an anticlockwise direction with reference to an arbitrary datum line or reference vector, which is usually drawn horizontally to the right of the centre of rotation as shown in fig. 7. After any time t seconds, the instantaneous value of the voltage, v , is shown by the height of the vertical projection of the end of the vector. In fig. 7 the vector has rotated through an angle of ωt

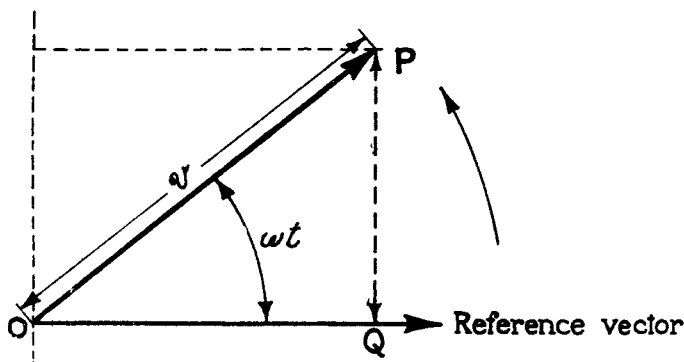


FIG. 7, CHAP. V.—Rotating vector.

radians, and the instantaneous voltage is shown to the same scale as the vector $\mathcal{V}$ by the line PQ which is equivalent to the projection of its length upon the vertical axis. In the study of alternating currents the advantage of vector representation is its ability to depict directly the conception of phase difference. When two such quantities of the same frequency pass through corresponding points in the wave-form at the same instant, they are said to be in phase with each other, while if they pass through corresponding points at different instants, there is said to be a phase difference between them, and one is said to be leading or lagging on the other. In order to show this a simple example may be taken, thus let $\mathcal{E}$ represent the peak value of an alternating E.M.F., and $\mathcal{I}$ the corresponding current, then if $\mathcal{E}$ and $\mathcal{I}$ are in phase they may be considered as vectors which are coincident in direction and rotate at equal speed. $\mathcal{E}$ and $\mathcal{I}$ are therefore represented as in fig. 8a. Actually $\mathcal{I}$ lies upon $\mathcal{E}$, but for clearness the two vectors are drawn side by side. In practice $\mathcal{E}$ may either be in phase with, or lead or lag upon

the resulting current ; in fig. 8b $\mathcal{I}$ is shown as leading, and in fig. 8c as lagging, with reference to the E.M.F. By convention the direction of rotation is always anticlockwise, and is shown on the vector diagram by a curved arrow.

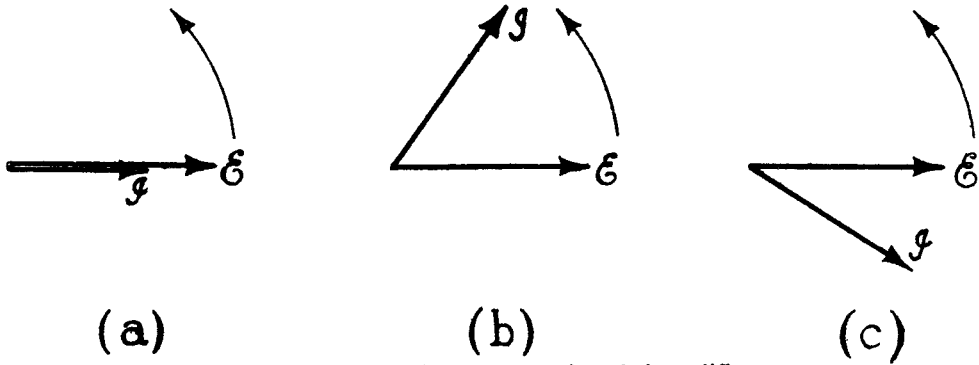


FIG. 8, CHAP. V.—Vector representation of phase difference.

Although the principle of vector representation is based upon the peak value, vector diagrams may be drawn to a scale of R.M.S. values, but it must then be remembered that the projection of the vector upon the vertical does not give the instantaneous value, although the phase difference between various quantities of the same frequency is still shown.

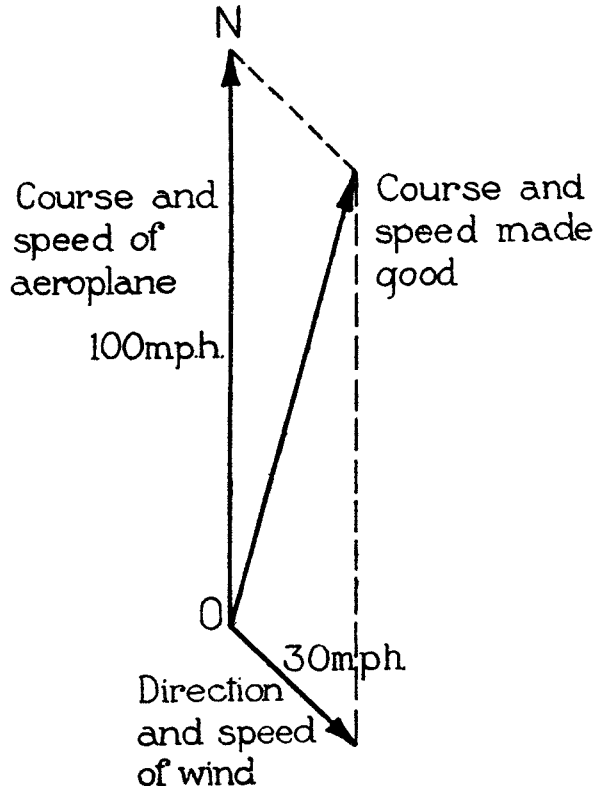


FIG. 9, CHAP. V.—Addition of vectors

CHAPTER V.—PARAS. 9-10

9. Like scalar or arithmetical quantities vectors may be added and subtracted. For instance, an aeroplane may have an actual speed of 100 miles per hour in a direction true North, but a wind from the North-west of 30 miles an hour may also act upon the machine. The velocity of the aeroplane with reference to the ground will then be the vector sum of the two velocities, and may be found by drawing the two vectors to scale, as shown in fig. 9. The two vectors form two sides of a parallelogram, and the vector sum can be found by completing the parallelogram (as shown by dotted lines) and then drawing the diagonal through the origin. The latter then represents the sum of the two vectors in magnitude and direction. Subtraction of vectors is carried out

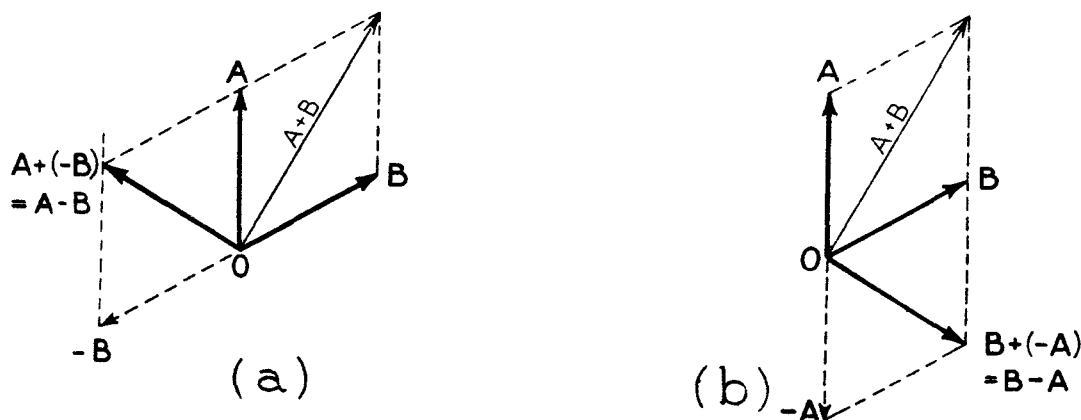


FIG. 10, CHAP. V.—Subtraction of vectors.

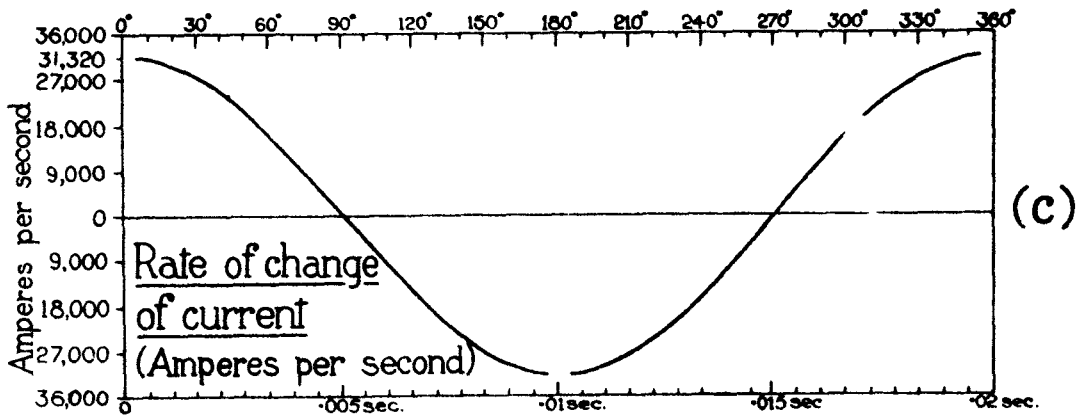
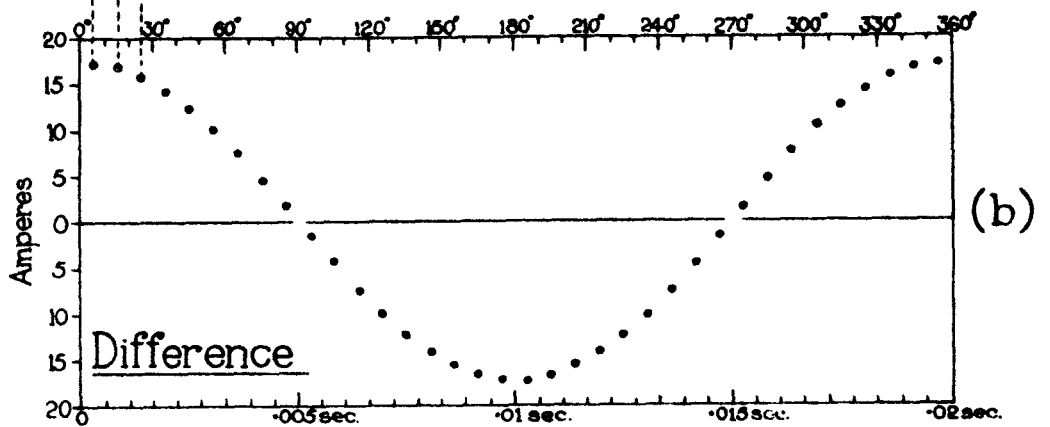
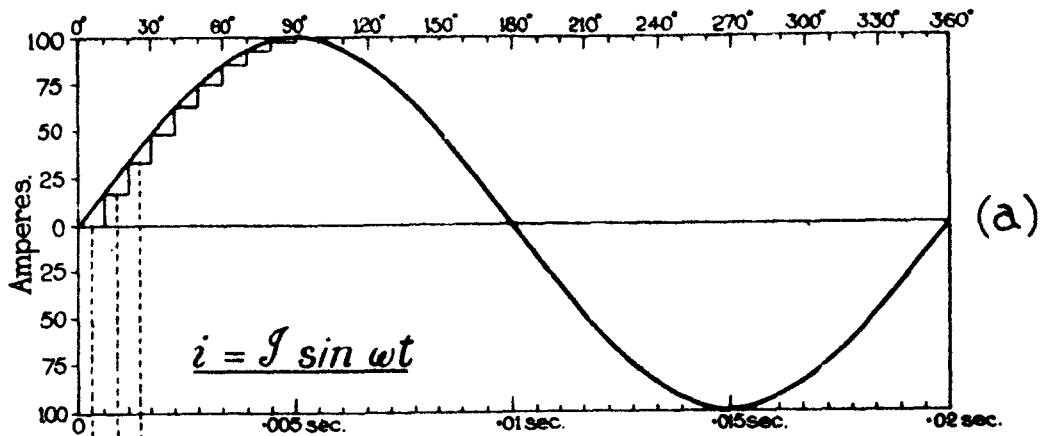
by the following method. Suppose that A and B are two vectors, and the vector difference $A - B$ is required. This may be written $A + (-B)$, and this expression gives the clue to the graphical solution, which is as follows. Draw the vectors A and B, then draw a vector equal in magnitude to B, but contrary in direction. This vector is then equal to $-B$. Perform the addition of the vectors A and $-B$ by the parallelogram rule given above. The diagonal of this parallelogram passing through the origin is equal to the vector $A - B$, and it must be clearly understood that the vector $A - B$ is not equal to the vector $B - A$, because its direction is different, as will be understood by reference to fig. 10.

The rate of change of a sinusoidal quantity

10. In earlier chapters, several illustrations of Faraday's law have been met, and it will be remembered that in its mathematical form the law is

$$e = L \frac{di}{dt}$$

Hitherto the symbol $\frac{di}{dt}$ has been considered merely as an abbreviation of the phrase "the rate of change of current with respect to time". The unit of $\frac{di}{dt}$ is the ampere per second, and its numerical value must be known before the self-induced E.M.F. can be calculated. Since an alternating current is constantly varying in magnitude, it becomes necessary to evaluate this quantity for a sinusoidal wave, and in order to exhibit the principle of the method a definite example may be taken. Fig. 11 (a) shows a sinusoidal current having a peak value of 100 amperes, and it is required to find the rate of change at every instant during the cycle. The period is $\cdot 02$ second, and this has been divided into 36 equal time intervals, each of $\frac{1}{1800}$ second; since one cycle is equal to 360° each interval corresponds to an angle of 10° . The instantaneous value



GRAPHICAL DERIVATION OF APPROXIMATE VALUE
OF $\frac{di}{dt}$ WHEN $i = I \sin \omega t$

of the current at each of these intervals in the first quarter of a cycle is given in the table below, and these values will be repeated, in the reverse order, during the second quarter of the cycle. Between 180° and 360° the current values will be a repetition of those in the first half-cycle but with negative sign.

Angle	..	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°
i	..	0	17.4	34.2	50	64.3	76.6	86.6	94	98.5	100
Difference	..	17.4	16.8	15.8	14.3	12.3	10	7.4	4.5	1.5	

In the third line of the table the change of current during each interval is given. During the first interval, the current increases from zero to 17.4 amperes, during the second from 17.4 to 34.2, an increase of 16.8 amperes, and so on. These successive increments of current are plotted in fig. 11b.

The average rate of change of current during each interval is found by dividing the change of current by the duration of the interval. For example during the first 10° the average rate of change is $17.4 \div \frac{1}{1800}$ or 31320 amperes per second. During the next 10° the average rate of change is $16.8 \div \frac{1}{1800}$ or 30240 amperes per second. On repeating this process for the whole of the first quarter of a cycle it will be seen that the rate of change of current is greatest where the current itself is small, but as the current increases the rate of change decreases. At the end of the first quarter of a cycle, the current reaches its peak value and is momentarily neither increasing nor decreasing. Its rate of change is therefore zero, and is so plotted in fig. 11b.

During the second quarter of a cycle the current is decreasing in value and its rate of change must be regarded as negative in sign, while during the third quarter the current is increasing in value but is of negative sign, and its rate of change must be regarded as negative. During the fourth quarter the current is decreasing in value and is still of negative sign so that the rate of change is again positive. The average rate of change of current has been plotted in fig. 11c and a smooth curve drawn through the points. This curve is incomplete because no values have been obtained for the rate of change at the instants $t = 0$, $t = .01$ second and $t = .02$ second. It can be seen nevertheless that it is very nearly a sine curve moved through a quarter of a cycle, that is, a cosine curve. If the period were divided into a larger number of intervals before carrying out the above process, the approximation would be closer still. It may now be stated that the instantaneous value of the rate of change of a sinusoidal current, $i = I \sin \omega t$ is a co-sinusoidal quantity and may be represented by the equation

$$\frac{di}{dt} = K \cos \omega t$$

where K is the maximum rate of change. The maximum rate of change occurs when the current itself is passing through the value zero, but its exact value cannot be obtained by the method used above. It can however be deduced as follows. Referring to fig. 12, let θ be the magnitude of the angle ROP in radians, and the length of the radius OR be r . Since the radian measure of an angle is given by the ratio $\frac{\text{arc}}{\text{radius}}$, $\theta = \frac{\text{arc RP}}{r}$, while $\sin \theta = \frac{QP}{r}$. It is obvious that the arc

RP is of greater length than the perpendicular QP, and therefore that $\frac{\sin \theta}{\theta}$ is less than unity. Now allow the angle θ to decrease to θ' by rotating OP into the position OP'. The magnitude of θ' is given by the ratio $\frac{\text{arc RP}'}{r}$ while $\sin \theta' = \frac{QP'}{r}$ and $\frac{\sin \theta'}{\theta'}$ although smaller than unity has approached nearer to that value. In the lower diagram of fig. 12 the angle θ has been made still smaller; the arc RP and perpendicular QP are now very nearly of the same length, and it

CHAPTER V.—PARA. 11

is deduced that if the angle decreases without limit, or approaches the value zero, the ratio $\frac{\sin \theta}{\theta}$ approaches the value unity. This conception is applied in finding the peak value of the rate of change of current by taking a small interval of time, δt , measured from the time $t = 0$. As $i = \mathcal{I} \sin \omega t$, at the end of the interval δt the current will be equal to $\mathcal{I} \sin \omega \delta t$ and the average rate of change $\frac{\delta i}{\delta t}$ during the interval will be $\frac{\mathcal{I} \sin \omega \delta t}{\delta t}$. As the interval δt is made smaller and smaller so the expression $\mathcal{I} \sin \omega \delta t$ becomes more nearly equal to $\mathcal{I} \omega \delta t$ and

$$\frac{\delta i}{\delta t} = \frac{\mathcal{I} \omega \delta t}{\delta t} = \omega \mathcal{I}$$

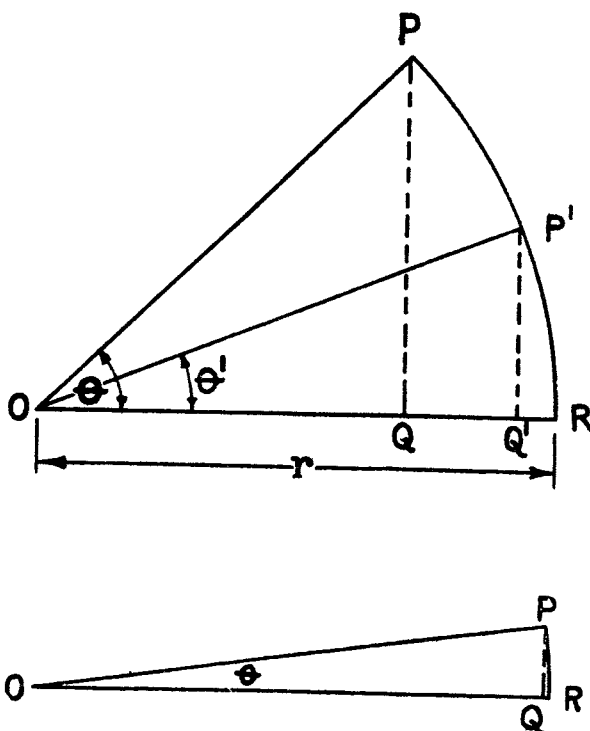


FIG. 12, CHAP. V.—Relation between arc θ and $\sin \theta$.

This is the maximum value of the rate of change of current. Hence, if

$$i = \mathcal{I} \sin \omega t$$

$$\frac{di}{dt} = \omega \mathcal{I} \cos \omega t$$

In the above example, $f = \frac{1}{T} = 50$ cycles per second and the maximum rate of change of current is $2\pi \times 50 \times 100$ or 31416 amperes per second.

11. It is often necessary in alternating current work to find the rate of change of a quantity. Let us first consider a simple instance of varying velocity taking a motor car starting from rest as an example. On letting in the clutch, the car starts to move, with increasing velocity. Now the velocity is the rate at which its position is changing, while the

rate at which its velocity is increasing is the acceleration. Thus a certain car was found to increase its speed from 10 to 31 miles per hour in seven seconds, and its acceleration is $\frac{31-10}{7}$ miles per hour per second. or since 1 mile = 5280 feet, the acceleration is $\frac{21 \times 5280}{3600 \times 7}$, or 4.4 ft. per second per second. When the car has travelled x feet its velocity v is $\frac{dx}{dt}$, while its acceleration is the rate at which its velocity is changing, or $\frac{dv}{dt}$. Now as $v = \frac{dx}{dt}$, the acceleration is $\frac{d}{dt} \left(\frac{dx}{dt} \right)$ and as a matter of convenience this is written $\frac{d^2x}{dt^2}$. It must be fully realised that the indices are not "powers" to which the quantities have to be raised, e.g. d^2 is not "d-squared"; the symbol $\frac{d^2x}{dt^2}$ must be regarded merely as a kind of shorthand sign for "the acceleration of a body in the direction x ".

In the case of a sinusoidal quantity such as $i = \mathcal{I} \sin \omega t$ the rate of change of which is $\frac{di}{dt} = \omega \mathcal{I} \cos \omega t$, to find the rate of change of the latter it is necessary to repeat the procedure used in finding $\frac{di}{dt}$ but operating upon the cosine curve instead of the sine curve. The detailed process will not be repeated, but the results can be stated thus:—

$$\begin{aligned} \text{If } i &= \mathcal{I} \sin \omega t \\ \frac{di}{dt} &= \omega \mathcal{I} \cos \omega t \\ \frac{d^2i}{dt^2} &= -\omega^2 \mathcal{I} \sin \omega t = -\omega^2 i \end{aligned}$$

ALTERNATING CURRENT CIRCUITS

A.C. resistance

12. When energy is supplied to any electrical circuit, some portion may be expended in doing useful work, but a portion is always expended in producing heat. If a circuit has a resistance R and the R.M.S. current flowing through it is I amperes, energy is converted into heat at a rate of $I^2 R$ joules per second. In direct current circuits, the resistance of any conductor can be calculated from the relation $R = \frac{l \rho}{A}$. In alternating current circuits, however, the heating effect is not confined to the conductor, for any and every other conductor situated within the magnetic field set up by an alternating current is the seat of currents due to the induced E.M.F. (Faraday's law) and consequently energy is converted into heat in these conductors. If any ferro-magnetic material is situated in this field, the continual reversal of the direction of magnetisation necessitates a further expenditure of energy, giving rise to what is called the hysteresis loss. The rate at which energy is dissipated by induced currents and by hysteresis depends upon the frequency of the A.C. supply. Energy may also be dissipated in the form of radiation. The consideration of this phenomenon is deferred until a later chapter, but the position may be summarised in the statement that for alternating current practice a new definition of resistance must be introduced, namely: The total effective resistance of an A.C. circuit is that quantity which, when multiplied by the square of the R.M.S. current, is equal to the rate at which energy is dissipated. If P is the rate of energy dissipation in joules per second or watts,

$$P = I^2 R$$

$$\text{and } R = \frac{P}{I^2}$$

CHAPTER V.—PARA. 12

The total effective resistance includes that due to all causes of power dissipation. As copper is almost universally employed for electrical conductors the power loss due to the conductor itself is often referred to as the copper loss. This in turn may be divided into three components, (i) the D.C. resistance, given by the formula $R = \frac{l\rho}{A}$ (ii) the resistance due to a phenomenon called skin effect, which may be described as a tendency for the current to flow on the surface of the conductor instead of being uniformly distributed over its cross-section and (iii) that due to what is termed proximity effect, which causes the current to concentrate in those portions of the conductor which are most remote from other conductors. Both skin and proximity effects are caused by the induction of eddy currents. Fig. 13 shows a portion of a conductor carrying an alternating current which is represented by solid lines parallel to the axis of the wire. The current gives rise to an alternating magnetic field, both inside and outside the conductor, and therefore to induced E.M.F.'s which cause eddy currents to circulate in paths somewhat as shown by the dotted lines, in opposition to the main current in the centre of the wire but in the same direction at the surface. The effect of this non-uniform current distribution is to cause an increase in the effective resistance of the conductor. To understand

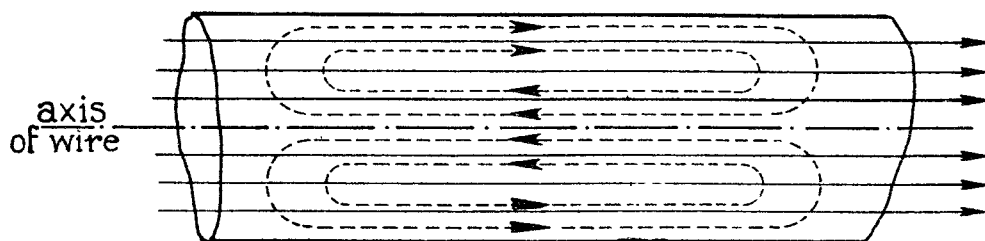


FIG. 13, CHAP. V.—Skin effect in isolated conductor.

how this increase occurs, imagine a conductor of D.C. resistance R ohms, of square cross-section and one inch side, to carry a current of I amperes. Divide the cross-section into four equal portions; each quarter of the conductor will then have a resistance of $4R$ ohms and with a uniform current distribution will carry $\frac{I}{4}$ amperes. The power dissipated in the conductor will therefore be $4 \times \left(\frac{I}{4}\right)^2 \times 4R = I^2 R$ watts. Now suppose that owing to some peculiar phenomenon, one of the quarters is prevented from carrying current, but that the total current remains as before, namely I amperes. Each remaining quarter must therefore carry $\frac{I}{3}$ amperes, and the power dissipated will be $3 \times \left(\frac{I}{3}\right)^2 \times 4R = \frac{4}{3} I^2 R$ watts.

By the definition given above the total effective resistance is the power dissipated divided by the square of the current, or $\frac{4}{3} I^2 R \div I^2 = \frac{4}{3} R$, and the A.C. resistance of the conductor in this particular instance is $\frac{4}{3}$ times the D.C. resistance. If several conductors are situated in close proximity, as in an inductively wound coil, each turn will be the seat of eddy currents set up by the current in adjacent turns as well as by the current in the turn itself. The current is then constrained to flow in paths having the form shown by the shaded areas in fig. 14, and the resistance is still further increased. This is the proximity effect mentioned above.

The total copper loss in a conductor may be obtained by adding the losses due to skin and proximity effects to the ordinary D.C. loss. Considering only the resistance of wires of circular

cross-section it can be shown that the increase of resistance is proportional to a factor $z = \pi d \sqrt{\frac{2f\mu}{\rho}}$ where d is the diameter of the wire in centimetres, f the frequency, μ the permeability of the material and ρ its specific resistance in E.M. units.

The A.C. resistance of a straight wire remote from all other conductors is

$$R_s = R_{dc} (1 + F)$$

where F is the skin effect factor. If z is less than 2, $F = \frac{z^4}{192}$, while if z is greater than 100, F approaches the value $\frac{\sqrt{2}z - 3}{4}$. Table VIII, Appendix A gives the value of F for the range $z = 0.1$ to $z = 100$.

The energy loss due to proximity effect may also be calculated and added to the other components to give the total energy loss. Taking the simplest case of two parallel wires of

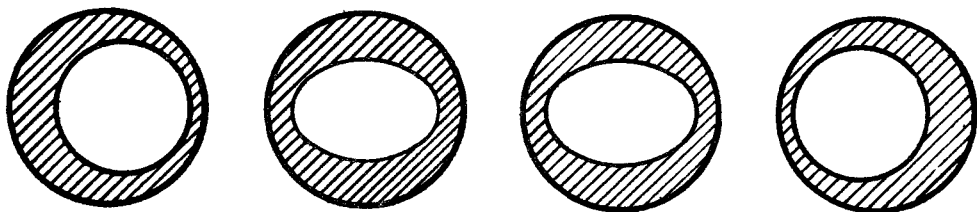


FIG. 14, CHAP. V.—Distribution of current in adjacent conductors (proximity effect).

diameter d centimetres the axes of which are separated by a distance of c centimetres, the additional resistance R_h due to proximity effect is

$$R_h = R_{dc} \frac{G d^4}{c^2}$$

where G , the proximity factor, also depends upon the value of z . If z is less than 0.5, $G = \frac{z^4}{64}$ approximately, while if z is greater than 100, $G = \frac{\sqrt{2}z - 1}{8}$. The total A.C. resistance then

becomes $R_{ac} = R_{dc} \left(1 + F + G \frac{d^4}{c^2} \right)$ provided that the ratio d/c does not approach unity closely. If however the wires are very close together a further correction must be applied in the form of a factor J and

$$R_{ac} = R_{dc} \left(1 + F + \frac{G \frac{d^4}{c^2}}{1 - J \frac{d^2}{c^2}} \right)$$

This formula may also be extended to apply to any number of spaced parallel wires, becoming

$$R_{ac} = R_{dc} \left(1 + F + \frac{u G \frac{d^4}{c^2}}{1 - J \frac{d^2}{c^2}} \right)$$

where u depends upon the number of wires in parallel. Values of G , J and u are given in Tables IX and X, Appendix A.

CHAPTER V.—PARAS. 13-14

13. The above expressions may be found to be of practical value in calculating the resistance of aerial wires and feeder lines, or of solenoids or flat spiral coils, provided they are wound in a single layer and that the coil radius is large compared with the winding length or depth. In other cases the calculation of the proximity effect becomes extremely complicated and only some practical conclusions can be given.

(i) Both the skin effect and proximity effect would be zero if equal current distribution could be achieved. An approximation to equal current distribution can be attained by using cable consisting of very fine strands of wire so plaited or twisted that every wire occupies successively a position near the centre and on the outside, each wire being thus of equal inductance and resistance. The strands may be made up in groups of three, similar to a rope, and it is found that there is an optimum value for the diameter of the constituent wires, depending upon the frequency. At extremely high frequencies it is impossible in practice to approach the extremely fine stranding demanded by theory, and a solid wire offers less resistance than a stranded one of incorrect design. To obtain any benefit from the use of "litz" cable, as this kind of wire is known, a very high standard of workmanship is desirable. A broken strand, or a single strand badly soldered to the terminal of the coil, may cause the latter to be less efficient than an ordinary coil of solid wire. In the repair of transmitting inductances and similar apparatus this should be constantly borne in mind.

(ii) If solid wire is used, e.g. for receiving inductances, there is an optimum gauge of wire for any given coil dimensions and frequency. As an example, consider the dimensions shown in fig. 15 which were chosen for a coil of 2,000 microhenries intended for reception of Droitwich

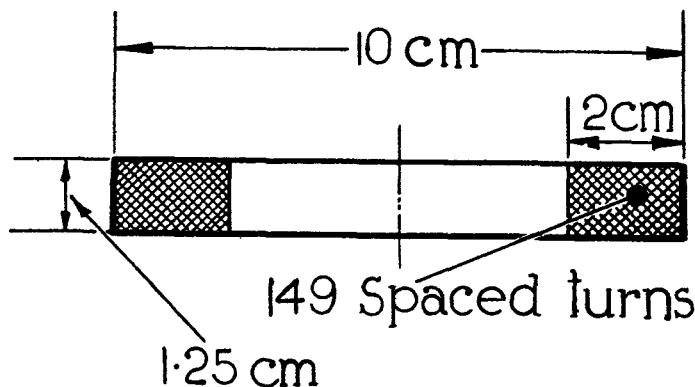


FIG. 15, CHAP. V.—Dimensions of multi-layer coil, 2,000 μ H.

(200 kc/s). Different sizes of wire were used for a number of windings, the winding space being completely filled on each occasion by spacing the turns as much as possible. It was found that if wound with the requisite number of turns of 36 s.w.g. its total resistance at the given frequency was 22.9 ohms. This was reduced to 15.6 ohms by rewinding with 32 s.w.g. but a further increase of wire diameter increased the resistance to 19.7 ohms. The greater portion of this increase is due to the proximity effect. An increase of wire gauge to 24 s.w.g. increased the resistance to 40 ohms.

(iii) Where heavy high-frequency currents are to be carried, copper tube is as efficient as solid copper, and if this is not available copper strip should be employed. Copper tube is the only conductor suitable for transmitting inductances for frequencies above 3,000 kc/s, and is preferably silver-plated, particularly if rubbing contacts are to be employed. The silver plating prevents oxidation and great care should be taken not to destroy it by harsh cleaning processes.

Resistance in an A.C. circuit

14. It is often convenient to consider the case where any of the three properties, inductance, capacitance or resistance, is entirely absent from a given circuit, for in practice the effects of one

or more of these is often negligible. Suppose therefore that a certain circuit is entirely devoid of capacitive and inductive effects, and consider the application of an E.M.F. $e = \mathcal{E} \sin \omega t$ to a resistance of R ohms. As there is no other E.M.F. (such as a counter-E.M.F. of self-induction) and no tendency for electrons to accumulate in any part of the circuit, i.e. no capacitance, a current will commence to flow as soon as the E.M.F. is applied, and this current will be proportional to the E.M.F. Thus in a circuit of this nature Ohm's law is obeyed, and the instantaneous value of the current is $\frac{e}{R}$ or $\frac{\mathcal{E} \sin \omega t}{R}$. The peak value of the current is $\mathcal{I} = \frac{\mathcal{E}}{R}$ and its effective or R.M.S. value $I = \frac{E}{R}$. The current will be in phase with the applied E.M.F. as shown by curves and vectors in fig. 16.

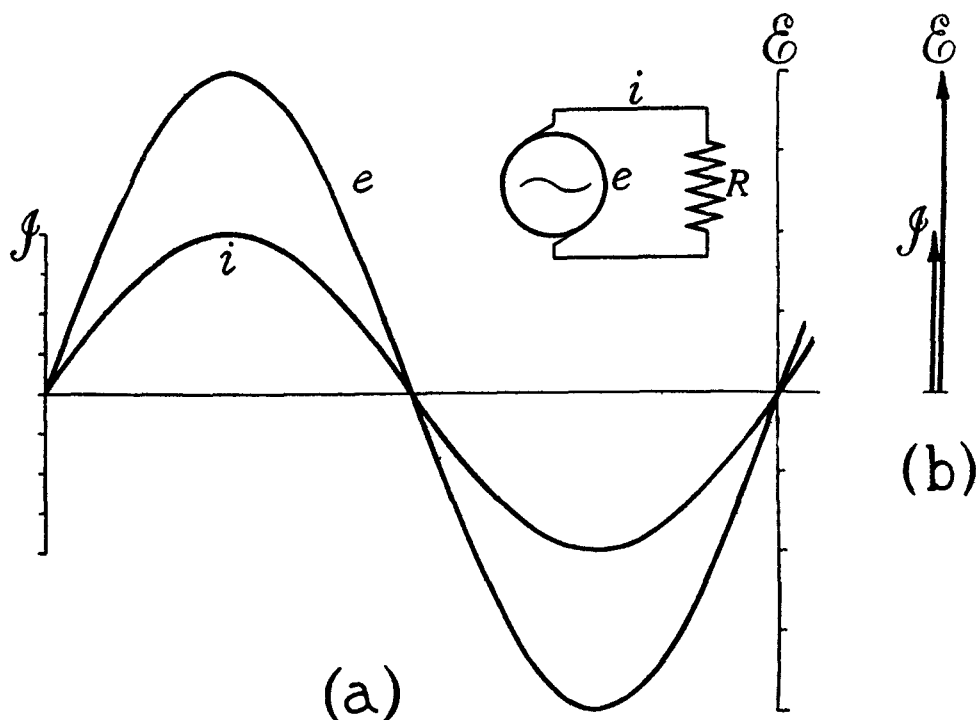


FIG. 16, CHAP. V.—Effect of resistance in A.C. circuit.

Inductance

15. The inductance of a given circuit is not truly a constant, but depends to some extent upon the frequency and intensity of the current flowing in it. This is particularly true if any ferro-magnetic substance is situated within the field of the coil. It has been shown that the value of the inductance is given by an equation having the form $L = \frac{KN^2}{S}$ where S is the reluctance of the path of the tubes of magnetic flux. When iron or any other ferro-magnetic material is present in the magnetic field, the reluctance varies with the factors above-mentioned, because both affect the effective permeability of the magnetic path. It is usual to distinguish between inductances designed for low frequencies, below about 10,000 cycles per second, and high frequencies which are those higher than 10,000 cycles per second. This dividing line between high and low frequencies cannot be drawn with precision and for some purposes it is desirable

CHAPTER V.—PARA. 15

to include frequencies up to about 20,000 cycles per second in the "low frequency" category. This division corresponds to the fact that vibrations of the air of from 16 up to about 20,000 cycles per second produce the sensation of sound, and it is usual to refer to this range of frequency as the audio-frequency band even if no question of sound production is involved. The frequencies above the audible limit are referred to in a corresponding manner as radio-frequencies even if no question of radiation arises.

Inductances intended for use in circuits carrying audio-frequency currents are invariably fitted with iron cores, which are laminated to reduce eddy current loss, while the iron used is of a kind which has low hysteresis loss. In the audio-frequency portions of radio apparatus, conditions frequently arise in which it is necessary that the winding should carry a current of complex wave-form, consisting of a steady component and several alternating components. The inductance of such a coil under these circumstances depends upon the magnitudes of both the A.C. and D.C. components. To illustrate this, consider the magnetisation or B/H curve, shown in fig. 17. Suppose that the magnetising force due to the D.C. is that corresponding to

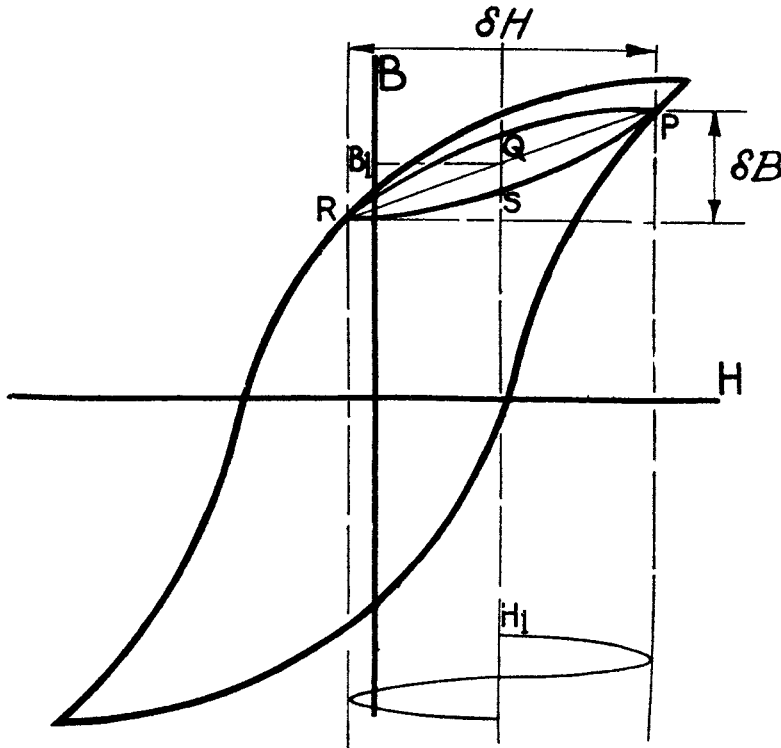


FIG. 17, CHAP. V.—Hysteresis loop showing definition of incremental permeability.

the ordinate H_1 , then the average flux density in the core will be B_1 . The magnetising force due to the A.C. will be superimposed upon this, and the iron will be carried through a cycle of magnetisation, as shown by the small hysteresis loop SPQR. The effective permeability with respect to the alternating current is then not the ratio B_1/H_1 as for direct current, but the average slope of the small loop, which may be represented by $\frac{\delta B}{\delta H}$, and this is always smaller than B_1/H_1 ,

hence the inductance of the coil is reduced by the presence of the direct current. The larger the steady magnetising force the less will be the value of $\frac{\delta B}{\delta H}$. The expression $\frac{\delta B}{\delta H}$ is called the incremental permeability to distinguish it from the ordinary permeability B/H . The introduction of an air gap into the iron core results in an increase in reluctance of the magnetic path, but also an increase in the incremental permeability. The net effect may be an increase in the effective inductance for alternating current while the inductance will also tend to remain constant, hence the cores of inductances for use under conditions in which the conductor carries both D.C. and A.C. almost invariably contain a small air gap.

16. Except in special circumstances, iron cores are not used in inductances designed for radio-frequency circuits for two reasons. First, the eddy current loss can only be kept within reasonable limits by sub-division of the core material (i.e. lamination) to a degree which is often impracticable. Second, it is desirable that the value of the inductance shall be independent of both the frequency and magnitude of the current. Radio-frequency circuits therefore usually employ air-core inductances, typical specimens of which have been briefly described in Chapter II. The capacitance which exists between all adjacent conductors must also be taken into account, for it is obvious that the presence of any metal inside the coil will increase this capacitance, whereas it is usually desirable to maintain it at the lowest possible value.

In spite of these difficulties, however, the cores of certain inductances in radio-frequency circuits are of iron or nickel-iron alloy in the form of a very fine powder which is amalgamated with a phenolic material as a binder. These coils are largely used in aircraft receivers, and receive further mention in a later chapter.

Inductance in an A.C. circuit

17. The effect of inductance in any circuit is to oppose any change in the value of the current, owing to the counter-E.M.F. set up by the changing flux linkage. If the inductance of the circuit is L henries, Faraday's and Lenz's laws tell us that the counter-E.M.F. is $e_b = -L \frac{di}{dt}$ volts. Now let us assume that an alternating current $i = \mathcal{I} \sin \omega t$ flows in a circuit having an inductance of L henries, but of negligible resistance and capacitance. Earlier in this chapter it was shown that if the current is of the assumed form, the rate of change of current is $\frac{di}{dt} = \omega \mathcal{I} \cos \omega t$. The counter-E.M.F. of self induction is therefore $-\omega L \mathcal{I} \cos \omega t$.

The applied E.M.F. must be equal and opposite to this counter-E.M.F. and no more, since the only work which the E.M.F. has to perform is to maintain a sinusoidal flux in the inductance. The applied E.M.F. is therefore

$$\begin{aligned} e &= \omega L \mathcal{I} \cos \omega t \\ &= \omega L \mathcal{I} \sin \left(\omega t + \frac{\pi}{2} \right) \end{aligned}$$

Hence the applied E.M.F. is also sinusoidal in form, but leads on the current by $\frac{\pi}{2}$ radians or 90 degrees. The frequency of the applied E.M.F. is the same as that of the current, which is of course to be expected. The peak value $\mathcal{E}$ of the E.M.F. is $\omega L \mathcal{I}$ volts, and the phase relationship between $\mathcal{E}$ and $\mathcal{I}$ is shown by curves and vectors in fig. 18. It will be observed that this relation may be expressed either as " $\mathcal{E}$ leading on $\mathcal{I}$ " or " $\mathcal{I}$ lagging behind $\mathcal{E}$ ". The latter is more usual, and the relation between the current and voltage may be written

$$i = \frac{\mathcal{E}}{\omega L} \sin \left(\omega t - \frac{\pi}{2} \right)$$

CHAPTER V.—PARA. 18

It is often convenient to consider R.M.S. rather than instantaneous values. Since $I = \frac{\mathcal{I}}{\sqrt{2}}$ and $E = \frac{\mathcal{E}}{\sqrt{2}}$, the relation between I and E is given by the equation $I = \frac{E}{\omega L}$. Comparing this with Ohm's law for direct current $I = \frac{E}{R}$, it will be seen that ωL is analogous with R in deciding the magnitude of the current for a given applied voltage.

Although L itself is in henries, the expression ωL is in ohms, because the dimensions or physical attributes of the ratio of voltage to current must always be the same. The factor ωL is termed inductive reactance of the circuit, and may be thought of as the opposition offered by an inductance to the flow of an alternating current, but it must be borne in mind that in addition to limiting the magnitude of the current, it causes the current to lag behind the E.M.F. which produces it. The symbol for reactance is X , and inductive reactance is denoted by X_L thus, $X_L = \omega L$.

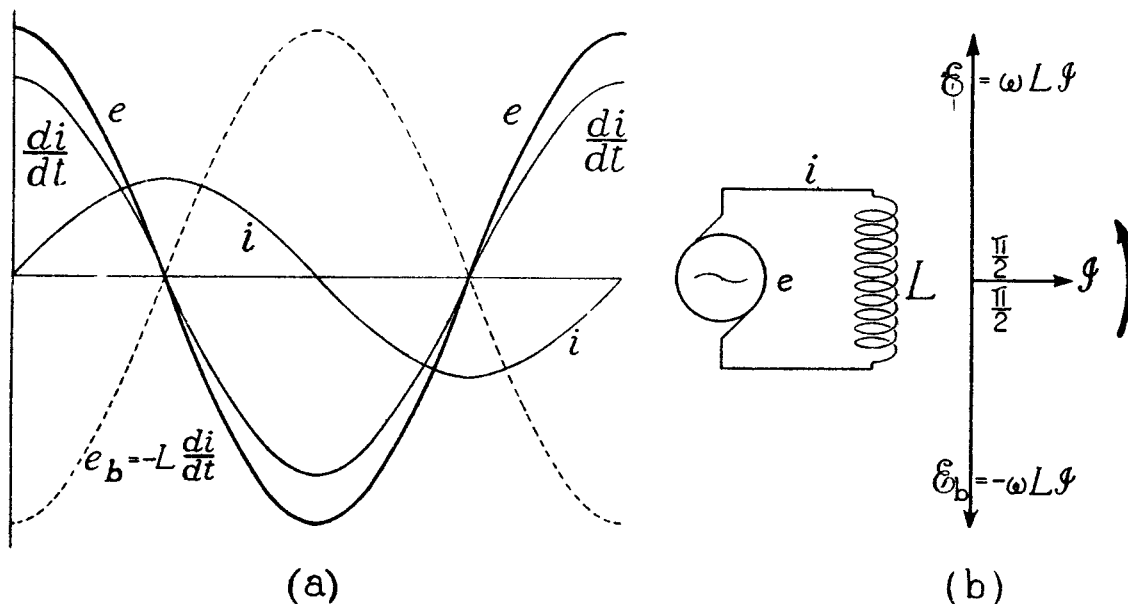


FIG. 18, CHAP. V.—Effect of inductance in A.C. circuit.

Inductances in series and parallel

18. If the alternating current $i = \mathcal{I} \sin \omega t$ flows through several inductances L_1, L_2, L_3 , etc., in succession, a counter-E.M.F. is set up in each inductance, equal to $-\omega L_1 i, -\omega L_2 i, -\omega L_3 i$, etc. The applied voltage must be sufficient to overcome the sum of all these counter-E.M.F.'s. and therefore, considering R.M.S. values only,

$$\begin{aligned} E &= \omega L_1 I + \omega L_2 I + \omega L_3 I, \text{ etc.} \\ &= \omega I (L_1 + L_2 + L_3 \dots \dots \dots) \\ &= \omega L I \end{aligned}$$

$$\text{where } L = L_1 + L_2 + L_3.$$

Hence the total inductance L is the sum of the individual inductances which are in series in the circuit.

If the inductances L_1, L_2, L_3 , etc., are placed in parallel, having a common sinusoidal P.D. of V volts between their ends, the current in each inductance, taking R.M.S. values as before, will be

$$I_1 = \frac{V}{\omega L_1}$$

$$I_2 = \frac{V}{\omega L_2}$$

$$I_3 = \frac{V}{\omega L_3}$$

$$I_1 + I_2 + I_3 = \frac{V}{\omega} \left(\frac{1}{L_1} + \frac{1}{L_2} + \frac{1}{L_3} \right)$$

$I_1 + I_2 + I_3$ is the total current flowing in the main or unbranched portion of the circuit and may be denoted by I . I and V are connected by the relation $I = \frac{V}{\omega L}$, where L is the joint inductance of L_1, L_2, L_3 in parallel.

Collecting these equations we see that

$$I = \frac{V}{\omega L}$$

$$I_1 + I_2 + I_3 = \frac{V}{\omega} \left(\frac{1}{L_1} + \frac{1}{L_2} + \frac{1}{L_3} \right)$$

$$\text{since } I = I_1 + I_2 + I_3$$

$$\frac{1}{L} = \frac{1}{L_1} + \frac{1}{L_2} + \frac{1}{L_3}$$

The reciprocal of the joint inductance is equal to the sum of the reciprocals of the individual inductances. It will be observed that the rules for inductances in series and in parallel are the same as for resistances.

A particular instance which often arises in practice is the calculation of the joint inductance of two inductances L_1 and L_2 in parallel. Since

$$\frac{1}{L} = \frac{1}{L_1} + \frac{1}{L_2}$$

giving the right-hand member of the equation a common denominator

$$\frac{1}{L} = \frac{L_1 + L_2}{L_1 L_2}$$

$$\text{whence } L = \frac{L_1 L_2}{L_1 + L_2}$$

and it is seen that the joint inductance of two inductances in parallel is given by the product of the two divided by their sum.

Effect of mutual inductance between coils in parallel

19. Consider the circuit given in fig. 19 which may represent the two coils of a variometer inductance. If an R.M.S. voltage E of frequency $\frac{\omega}{2\pi}$ is applied, a current of I_1 amperes will flow in the coil L_1 and a current of I_2 amperes in the coil L_2 . If the mutual inductance between the coils is zero the counter-E.M.F. induced in the two windings will be $-\omega L_1 I_1$ and $-\omega L_2 I_2$ volts respectively. If the mutual inductance has the finite value M henries, however, the

CHAPTER V.—PARA. 19

counter-E.M.F. induced in the coil L_1 will be $-(\omega L_1 I_1 + \omega M I_2)$ volts and in the coil L_2 , $-(\omega L_2 I_2 + \omega M I_1)$ volts. As the resistance is assumed to be negligible, the counter-E.M.F. is equal and opposite to the applied E.M.F. and

$$\omega L_1 I_1 + \omega M I_2 = E \quad \dots \quad (a)$$

$$\omega M I_1 + \omega L_2 I_2 = E \quad \dots \quad (b)$$

$$\text{or } \omega L_1 L_2 I_1 + \omega M L_2 I_2 = L_2 E \quad \dots \quad (c)$$

$$\omega M^2 I_1 + \omega M L_2 I_2 = M E \quad \dots \quad (d)$$

Subtracting (d) from (c)

$$\omega(L_1 L_2 - M^2) I_1 = (L_2 - M) E \quad \dots \quad (e)$$

$$\text{also } \omega L_1 M I_1 + \omega M^2 I_2 = M E \quad \dots \quad (f)$$

$$\omega L_1 M I_1 + \omega L_1 L_2 I_2 = L_1 E \quad \dots \quad (g)$$

Subtracting (f) from (g)

$$\omega(L_1 L_2 - M^2) I_2 = (L_1 - M) E \quad \dots \quad (h)$$

From (e) and (h)

$$I_1 = \frac{(L_2 - M) E}{\omega(L_1 L_2 - M^2)}; \quad I_2 = \frac{(L_1 - M) E}{\omega(L_1 L_2 - M^2)}$$

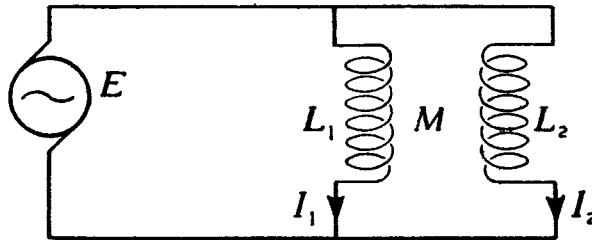


FIG. 19, CHAP. V.—Inductances in parallel, possessing mutual inductance.

The total current I is $I_1 + I_2$

$$I = \frac{L_1 + L_2 - 2M}{\omega(L_1 L_2 - M^2)} E$$

The effective reactance of the two coils in parallel is $\frac{E}{I} = \omega L$, and

$$\omega L = \frac{\omega(L_1 L_2 - M^2)}{L_1 + L_2 - 2M}$$

Hence the effective inductance of the parallel combination, including the mutual inductance, is

$$L = \frac{L_1 L_2 - M^2}{L_1 + L_2 - 2M}$$

When the coils are perpendicular to each other the mutual inductance is zero and the effective inductance is $\frac{L_1 L_2}{L_1 + L_2}$ as already shown. With any other relative disposition the mutual inductance has a finite value which may be either positive or negative. These signs are purely conventional, and it is convenient to regard the mutual inductance as positive if its effect is to increase the total inductance; this was adopted in writing the equations (a) and (b). The opposite signs are sometimes adopted in certain theoretical work. The effect of the mutual upon the total effective inductance is seen in the following example.

Example 3.—(i) The two coils of a variometer, each having an inductance of 100 microhenries, are connected in parallel. When the coils are co-axial, the mutual inductance is 90 microhenries. Find the maximum and minimum inductance.

$$\begin{aligned}
 \text{When } M \text{ is positive, } L &= \frac{L_1 L_2 - M^2}{L_1 + L_2 - 2M} \\
 &= \frac{L_1^2 - M^2}{2(L_1 - M)} \\
 &= \frac{(L_1 + M)(L_1 - M)}{2(L_1 - M)} \\
 &= \frac{L_1 + M}{2} = \frac{190}{2} = 95 \mu H
 \end{aligned}$$

$$\begin{aligned}
 \text{When } M \text{ is negative, } L &= \frac{(L_1 + M)(L_2 - M)}{2(L_1 + M)} \\
 &= \frac{L_1 - M}{2} = 5 \mu H
 \end{aligned}$$

(ii) If the mutual inductance were 98 microhenries, what would be the total range of inductance?

$$\text{From the above, with positive } M, L = \frac{100 + 98}{2} = 99 \mu H$$

$$\text{and with negative } M, L = \frac{100 - 98}{-2} = 1 \mu H$$

Hence the total range is from 99 to 1 microhenry. If every tube of magnetic flux linked with every turn of both coils, the inductance range would be from 100 to 0 microhenries.

Resistance and inductance in series

20. In the circuit diagram of fig. 20 a source of alternating E.M.F. supplies current to a circuit consisting of an inductance of L henries and a resistance of R ohms, connected in series. It is required to find the current which will flow, and the relative phases of current and voltage. The opposition offered by the circuit is now of two kinds (i) the resistance, which limits the value of the current, but will cause no phase difference, and (ii) the inductive reactance, which also limits the value of the current, and tends to cause it to lag behind the applied E.M.F. by $\frac{\pi}{2}$ radians. The applied E.M.F. can therefore be divided into two components, one of which may be considered to overcome the resistance, or to supply the energy converted into heat, and the other to overcome the inductive reactance or to supply energy which is stored in the form of a magnetic field when the current is increasing in value, and returned to the circuit when the current is decreasing.

The instantaneous values of these components may be denoted by v_R and v_L . Assuming the current to be sinusoidal, the curves in fig. 20 show the nature of their variation. The component v_R has the instantaneous value iR , and is in phase with the current, while the component v_L has the instantaneous value ωLi , and leads on the current by $\frac{\pi}{2}$ radians or 90° . The total voltage supplied by the alternator at any instant is found by adding the ordinates of the two curves, giving the resultant curve e , which represents the applied voltage. It is seen that the latter reaches its maximum value before the instant of maximum current, but that the angle of phase difference is less than 90° .

CHAPTER V.—PARA. 20

The relations between the maximum values $\mathcal{V}_R$, $\mathcal{V}_L$ and $\mathcal{E}$ are more rapidly obtained by a vector diagram (see fig. 20). The component $\mathcal{V}_R$ is in phase with $\mathcal{I}$ and the component $\mathcal{V}_L$ leads on $\mathcal{I}$ by 90° . The peak value of the applied E.M.F. is the vector sum of these and is obtained by the method previously described. Their sum is however easily obtained by the well-known "theorem of Pythagoras", i.e. if a and b are the two shorter sides of a right-angled triangle the length of the third side is $\sqrt{a^2 + b^2}$. Applying this theorem

$$\begin{aligned}\mathcal{E} &= \sqrt{\mathcal{V}_R^2 + \mathcal{V}_L^2} \\ \text{Now} \quad \mathcal{V}_R &= R \mathcal{I}, \mathcal{V}_L = \omega L \mathcal{I} \\ \text{and} \quad \mathcal{E} &= \sqrt{(R \mathcal{I})^2 + (\omega L \mathcal{I})^2} \\ &= \mathcal{I} \sqrt{R^2 + (\omega L)^2}\end{aligned}$$

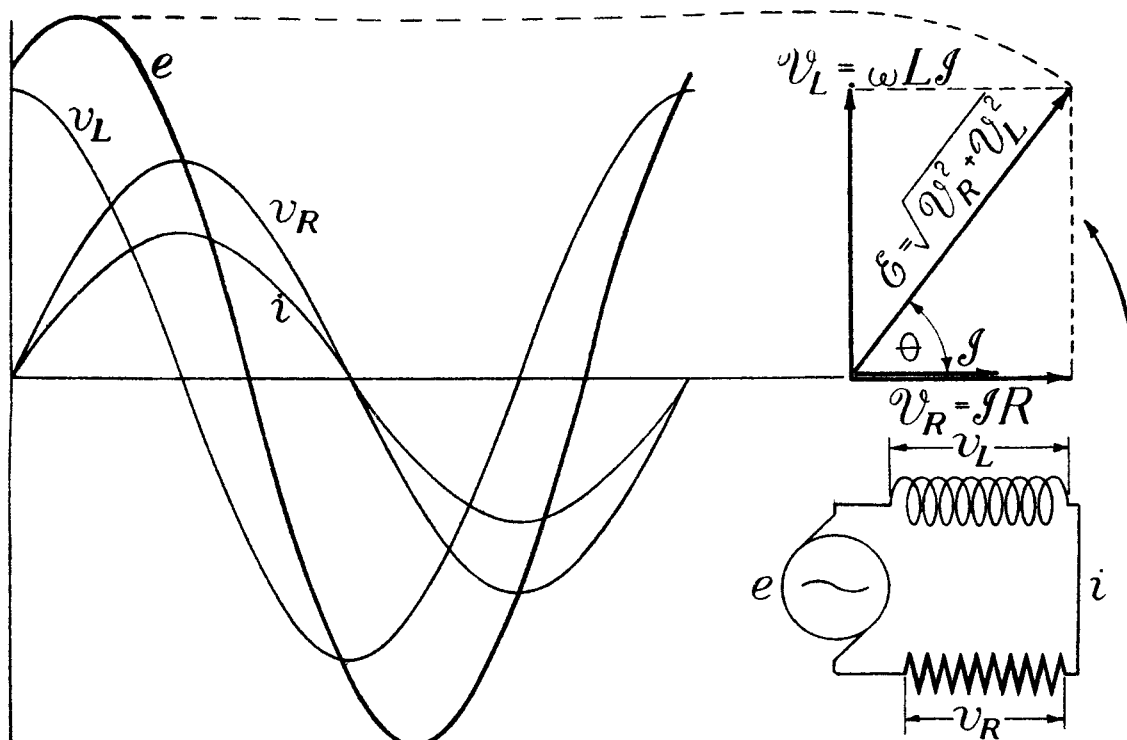


FIG. 20, CHAP. V.—Effect of inductance and resistance in series.

From the vector diagram it is obvious that the current lags upon the applied voltage by an angle θ . The magnitude of this angle is found from either of the following formulae:—

$$\begin{aligned}\tan \theta &= \frac{\omega L \mathcal{I}}{R \mathcal{I}} = \frac{\omega L}{R} , \\ \sin \theta &= \frac{\omega L}{\sqrt{R^2 + (\omega L)^2}} = \frac{\omega L}{Z} \\ \cos \theta &= \frac{R}{\sqrt{R^2 + (\omega L)^2}} = \frac{R}{Z}\end{aligned}$$

The instantaneous value of the current is

$$i = \frac{\mathcal{E}}{\sqrt{R^2 + (\omega L)^2}} \sin(\omega t - \theta)$$

where θ is the angle whose tangent is $\frac{\omega L}{R}$ (see above). The notation usually used is

$$\theta = \tan^{-1} \frac{\omega L}{R}$$

The R.M.S. value of the current is

$$I = \frac{E}{\sqrt{R^2 + (\omega L)^2}}$$

which may again be compared with Ohm's law. In this case $\frac{E}{I} = \sqrt{R^2 + (\omega L)^2}$, and $\sqrt{R^2 + (\omega L)^2}$ is in ohms. It may be considered to represent the total opposition of the current to the flow of current and is called the impedance of the circuit. The symbol Z is used for impedance when it is not necessary or possible to express it in a more detailed form.

Example 4.—An alternating E.M.F. of 220 volts peak value, having a frequency of 100 cycles per second, is applied to an inductance of 1.5 henries and a resistance of 600 ohms in series. Find (i) the R.M.S. value of the current; (ii) the peak P.D. at the terminals of the inductance; (iii) the angle of phase difference.

$$\omega = 2\pi f = 628$$

$$\omega L = 628 \times 1.5 = 942 \text{ ohms}$$

$$\begin{aligned} Z &= \sqrt{R^2 + (\omega L)^2} = \sqrt{600^2 + 942^2} \\ &= \sqrt{36 \times 10^4 + 88.75 \times 10^4} \\ &= 100 \sqrt{124.75} \\ &= 1120 \text{ ohms approx.} \end{aligned}$$

$$E = \frac{\mathcal{E}}{\sqrt{2}} = .707 \times 220 = 155.5 \text{ volts.}$$

$$(i) \quad I = \frac{E}{Z} = \frac{155.5}{1120} = .139 \text{ amperes (nearly).}$$

(ii) Peak P.D. at inductance terminals = $\omega L \mathcal{I}$ volts

$$\mathcal{I} = \sqrt{2} I = 1.414 \times .139 = 1.965 \text{ amperes}$$

$$\text{or} \quad \mathcal{I} = \frac{\mathcal{E}}{Z} = \frac{220}{1120} \text{ amperes} = 1.965 \text{ amperes}$$

$$\omega L \mathcal{I} = 942 \times 1.965 = 185 \text{ volts.}$$

$$\begin{aligned} (iii) \quad \theta &= \tan^{-1} \frac{\omega L}{R} \\ &= \tan^{-1} \frac{942}{600} \\ &= \tan^{-1} 1.57 \end{aligned}$$

Reference to a table of tangents gives $\theta = 58^\circ$, to the nearest degree, which is of sufficient accuracy for all practical work. Since the reactance is inductive, the current will lag on the applied E.M.F. by 58° .

CHAPTER V.—PARAS. 21-22

Impedances in series

21. If a number of pieces of apparatus having both resistance and inductance are connected in series and an alternating E.M.F. is applied, the resulting current, $\mathcal{I} \sin \omega t$, will set up a P.D. between the terminals of each instrument. If the resistance of each is denoted by R_1, R_2, R_3 , etc., and the corresponding inductance by L_1, L_2, L_3 , etc., the peak P.D.'s will be $\mathcal{V}_1 = \mathcal{I} \sqrt{R_1^2 + (\omega L_1)^2}$, $\mathcal{V}_2 = \mathcal{I} \sqrt{R_2^2 + (\omega L_2)^2}$, etc.

$\mathcal{I}$ will lag on $\mathcal{V}_1$, by an angle $\tan^{-1} \frac{\omega L_1}{R_1}$, on $\mathcal{V}_2$ by an angle $\tan^{-1} \frac{\omega L_2}{R_2}$, etc. The peak value, $\mathcal{E}$, of the applied E.M.F. will be equal to the vector sum of $\mathcal{V}_1, \mathcal{V}_2, \mathcal{V}_3$, and the resulting vector diagram is shown in fig. 21. It will be seen that

$$\mathcal{E} = \sqrt{\mathcal{I}^2 (R_1 + R_2 + R_3)^2 + \mathcal{I}^2 (\omega L_1 + \omega L_2 + \omega L_3)^2}$$

$$\text{or } \mathcal{E} = \mathcal{I} \sqrt{(R_1 + R_2 + R_3)^2 + \omega^2 (L_1 + L_2 + L_3)^2}$$

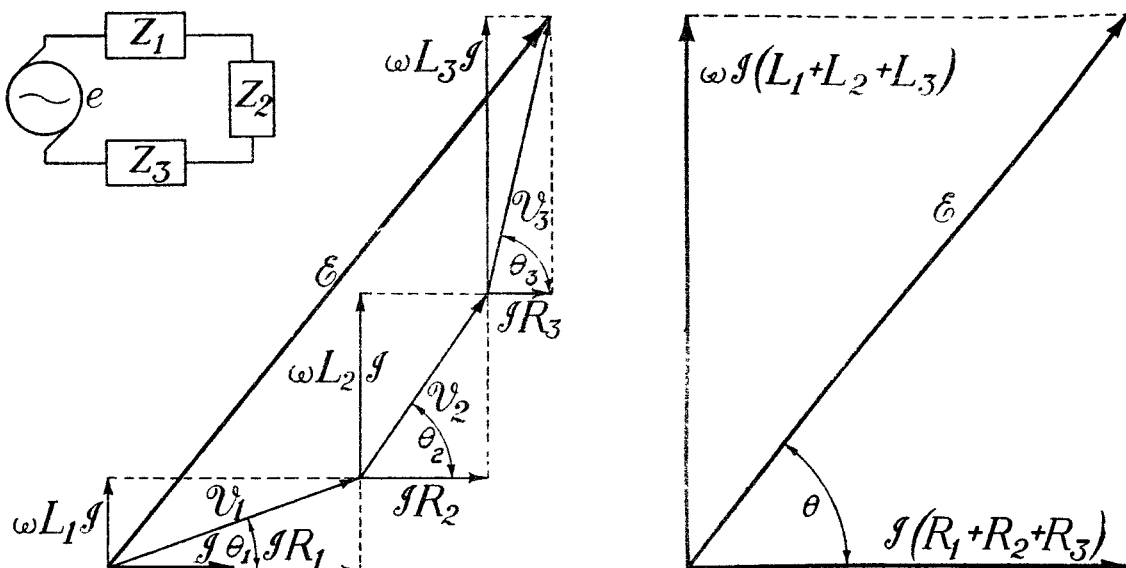


FIG. 21, CHAP. V.—Effect of inductive impedances in series.

The current $\mathcal{I}$ will lag on the applied E.M.F. $\mathcal{E}$ by an angle $\tan^{-1} \frac{\omega(L_1 + L_2 + L_3)}{R_1 + R_2 + R_3}$, and this is true no matter how many separate impedances of this kind are connected in series.

In any of the foregoing vector diagrams, the three vectors form the sides of a right-angled triangle having sides $\mathcal{I}Z, \mathcal{I}X, \mathcal{I}R$. The relative lengths of these sides, that is the shape of the triangle, will not be altered if each side is divided by $\mathcal{I}$, giving a triangle of sides Z, X, R . This is called the impedance triangle of the circuit and may be used instead of the true vector diagram, in order to obtain the angle of phase difference θ , by means of the equations $\tan \theta = \frac{X}{R}$,

$$\sin \theta = \frac{X}{Z}, \cos \theta = \frac{R}{Z}$$

Capacitance

22. In Chapter I the flow of electric current in a circuit was compared with the flow of water in a pipe line. Let us consider this analogy further. In fig. 22 the pipe line is supplied

from a pump which moves the water to and fro round the circuit instead of continuously in the same direction. The friction of the water against the sides of the pipe may be considered to represent the resistance of the electrical circuit, and the inertia of the water, that is its opposition to a change of motion, to represent its inductance. Suppose that the flow of water in the pipe line is restricted at one point by a flexible diaphragm, then provided that the pressure is insufficient

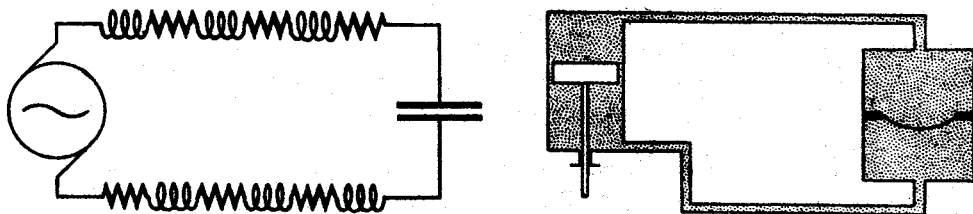


FIG. 22, CHAP. V.—Hydraulic analogue of A.C. circuit containing a condenser.

to burst this diaphragm, the pump can still move the water to and fro in the line, the diaphragm being stretched in one direction and the other alternately. Some work will be done by the pump in stretching the diaphragm, but this energy is stored in the diaphragm, and is expended in moving the water when the pressure of the pump is relaxed. To this extent therefore the diaphragm relieves the pump of an amount of work exactly equal to that which was expended in stretching it, and the action of the diaphragm is analogous to the presence of a condenser in an alternating current circuit.

Capacitance in an A.C. circuit

23. Let us now suppose that a condenser of capacitance C farads is connected directly to a source of alternating E.M.F. of peak value $\mathcal{E}$. This condenser will be presumed to have no energy losses and therefore no resistance, using the latter term in the extended sense applicable to A.C. theory. We have seen that the law connecting the capacitance C , the P.D., v , between its plates, and the charge q , is $q = Cv$, and this is applicable at all times, a change of P.D. being accompanied by a change of charge. If the condenser P.D. is caused by an applied voltage, the latter will always be equal to the P.D. but acting in the opposite direction, and as a result of this the condenser voltage is often referred to as the counter-E.M.F. of the condenser. The condenser voltage is not strictly an E.M.F. for by definition an E.M.F. only exists when energy of some other form undergoes conversion into electrical energy, and in the condenser the energy due to its charge is stored in electrical form.

When an alternating E.M.F. is applied to the condenser, the charge introduced into the latter will be directly proportional to the E.M.F. Hence if the applied E.M.F. is $e = \mathcal{E} \sin \omega t$ the instantaneous charge will be $q = C\mathcal{E} \sin \omega t$. The current flowing into, or out of the condenser is the rate at which electricity enters or leaves, and is measured in coulombs per second or amperes. This may be concisely expressed by saying that the current is the rate at which the charge is changing, and using the notation hitherto adopted for quantities which vary with time

$$i = \frac{dq}{dt}$$

Now $q = C\mathcal{E} \sin \omega t$ and from previous discussion it follows that the rate of change of the latter will follow a cosine law, hence

$$\begin{aligned} i &= \omega C \mathcal{E} \cos \omega t \\ &= \omega C \mathcal{E} \sin \left(\omega t + \frac{\pi}{2} \right) \end{aligned}$$

and it is apparent that the charging current varies in magnitude in a similar manner to the applied E.M.F. but leads on the latter by $\frac{\pi}{2}$ radians or 90° (fig. 23). If it is desired to express the

CHAPTER V.—PARA. 24

relationship between the E.M.F. and current in R.M.S. values, we may write $I = \omega CE$, which may again be compared with Ohm's law by rearranging in the form $I = E \frac{1}{\omega C}$. The denominator $\frac{1}{\omega C}$ which is analogous to the resistance in the true Ohm's law, is expressed in ohms, and is called the capacitive reactance of the condenser, the symbol X_c being sometimes used when it is unnecessary to introduce any reference to the frequency.

Condensers in series

24. Suppose that in a circuit to which is applied an E.M.F. $e = \mathcal{E} \sin \omega t$, we have a number of condensers C_1, C_2, C_3 , etc., arranged in series. Then the capacitive reactance or opposition of each will be $\frac{1}{\omega C_1}, \frac{1}{\omega C_2}, \frac{1}{\omega C_3}$ respectively, and the peak value of P.D. across each will consequently

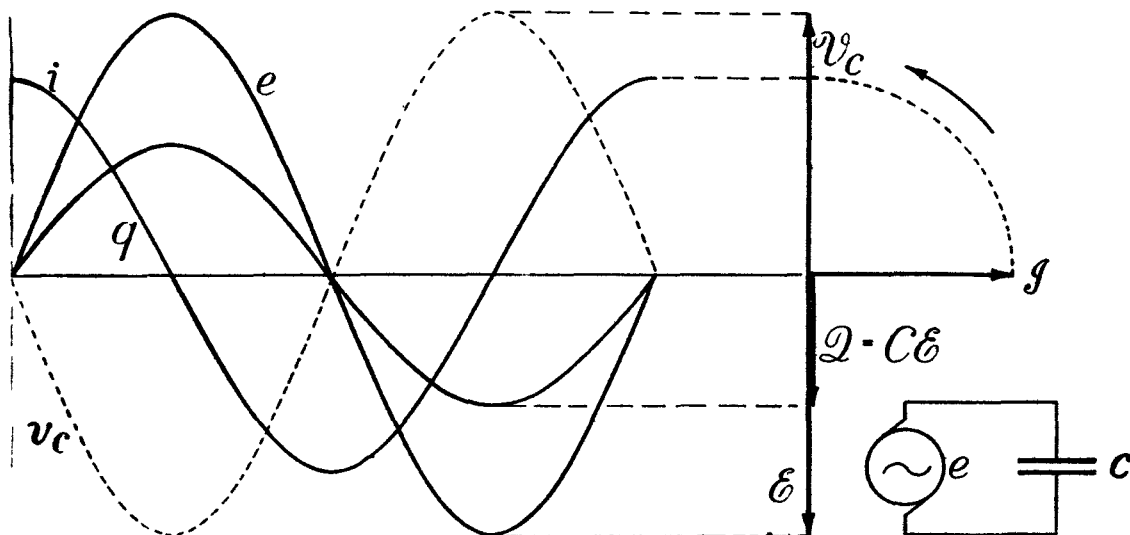


FIG. 23, CHAP. V.—Effect of capacitance in A.C. circuit.

be $\frac{\mathcal{E}}{\omega C_1}, \frac{\mathcal{E}}{\omega C_2}, \frac{\mathcal{E}}{\omega C_3}$. The total applied E.M.F. will be equal to the sum of the P.D's. Hence

$\mathcal{E} = \mathcal{E} \left(\frac{1}{\omega C_1} + \frac{1}{\omega C_2} + \frac{1}{\omega C_3} \right)$ But if C is the total effective capacitance of the circuit

$$\mathcal{E} = \frac{\mathcal{E}}{\omega C}$$

$$\text{and therefore } \frac{1}{C} = \frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3}$$

The effective capacitance of a number of condensers in series is therefore given by the reciprocal rule, just as for inductances or resistances in parallel.

When only two condensers are placed in series, a simpler formula is usually used.

$$\begin{aligned} \frac{1}{C} &= \frac{1}{C_1} + \frac{1}{C_2} \\ \frac{1}{C} &= \frac{C_2 + C_1}{C_1 C_2} \\ \therefore C &= \frac{C_1 C_2}{C_1 + C_2} \end{aligned}$$

Capacitance and resistance in series

25. Just as inductance and resistance may be found in series in an A.C. circuit, so a circuit may contain capacitance and resistance in series. The opposition offered by the circuit will now be of two kinds, (i) the resistance, which will limit the current without causing any phase difference between $\mathcal{E}$ and $\mathcal{I}$, and (ii) the opposition caused by the counter-E.M.F. due to the charge in the condenser, which limits the current and also tends to cause the current to lead on the E.M.F. by 90° . If a sinusoidal current is caused to flow in such a circuit the peak value of the applied E.M.F. must be equal to the vector sum of the two P.D's, (i) between the ends of the resistance R , $\mathcal{V}_R$, and (ii) between the terminals of the condenser C , $\mathcal{V}_C$, respectively. If the peak value of the current is $\mathcal{I}$, $\mathcal{V}_R = \mathcal{I}R$ and $\mathcal{V}_C = \frac{\mathcal{I}}{\omega C}$, while $\mathcal{E} = \sqrt{\mathcal{V}_R^2 + \mathcal{V}_C^2}$.

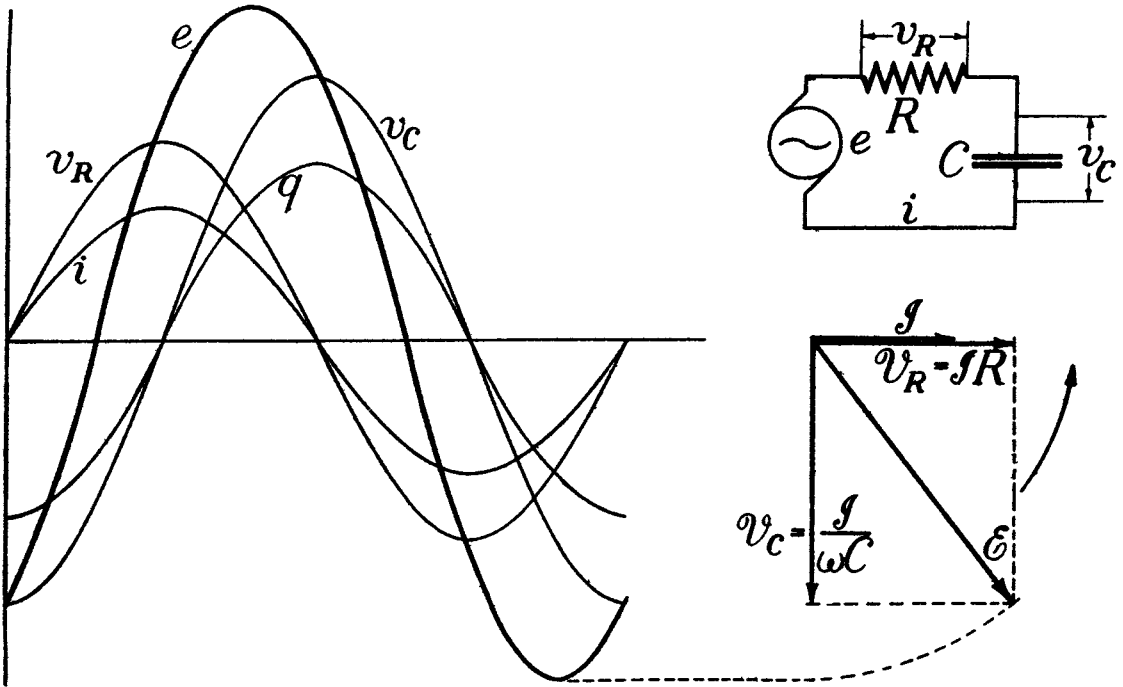


FIG. 24, CHAP. V.—Effect of capacitance and resistance in series.

$\mathcal{V}_R$ is in phase with $\mathcal{I}$, and $\mathcal{V}_C$ leads on $\mathcal{I}$ by 90° . The resulting phase relation between $\mathcal{E}$ and $\mathcal{I}$ is shown by the curves and vectors of fig. 24. With the aid of the vector diagram we deduce that

$$\mathcal{E} = \mathcal{I} \sqrt{R^2 + \left(\frac{1}{\omega C}\right)^2}$$

and that $\mathcal{I}$ leads on $\mathcal{E}$ by an angle θ ,

$$\text{where} \quad \theta = \tan^{-1} \frac{\frac{1}{\omega C}}{R} = \frac{1}{\omega CR}$$

The instantaneous value of the current is therefore

$$i = \frac{\mathcal{E}}{\sqrt{R^2 + \left(\frac{1}{\omega C}\right)^2}} \sin(\omega t + \theta)$$

CHAPTER V.—PARA. 26

In R.M.S. values, again

$$I = \frac{E}{\sqrt{R^2 + \left(\frac{1}{\omega C}\right)^2}},$$

and the expression

$$\sqrt{R^2 + \left(\frac{1}{\omega C}\right)^2} = \sqrt{R^2 + X_c^2} = Z$$

is the impedance of the circuit, in ohms.

Example 5.—A condenser of .0015 microfarads and a resistance of 20 ohms are connected in series and an alternating E.M.F. of 1.5 millivolts R.M.S. at a frequency of 10^6 cycles per second is applied to the circuit. Find (i) the R.M.S. value of the resulting current (ii) the peak P.D. at the condenser terminals and (iii) the angle of phase difference.

$$\omega = 2\pi f = 2\pi \times 10^6$$

$$\begin{aligned} X_c &= \frac{1}{\omega C} = \frac{1}{2\pi \times 10^6 \times .0015 \times 10^{-6}} \\ &= \frac{10^3}{2 \times 1.5} \\ &= \frac{10^3}{3} = 106 \text{ ohms} \end{aligned}$$

$$\begin{aligned} Z &= \sqrt{R^2 + \left(\frac{1}{\omega C}\right)^2} \\ &= \sqrt{20^2 + 106^2} \\ &\doteq 108 \text{ ohms.} \end{aligned}$$

$$\begin{aligned} \text{(i)} \quad I &= \frac{E}{Z} = \frac{1.5}{108} \text{ milliamperes} \\ &= .0139 \text{ milliamperes.} \end{aligned}$$

$$\text{(ii)} \quad \mathcal{V}_c = \frac{\mathcal{I}}{\omega C}, \quad \mathcal{I} = \sqrt{2} I = 1.414 \times .0139 = .01965 \text{ milliamperes}$$

$$\therefore \mathcal{V}_c = 106 \times .01965 = 2.08 \text{ millivolts.}$$

$$\begin{aligned} \text{(iii)} \quad \theta &= \tan^{-1} \frac{X_c}{R} \\ &= \tan^{-1} \frac{106}{20} \\ &= \tan^{-1} 5.3 \end{aligned}$$

$$\therefore \theta \doteq 80^\circ$$

i.e. $\mathcal{I}$ leads on $\mathcal{E}$ by 80° .

Inductance, capacitance and resistance in series

26. In fig. 25 is shown an alternating E.M.F. of peak value $\mathcal{E}$, applied to a circuit containing a resistance R , an inductance L , and a condenser C , in series. The alternator in this instance has to perform three duties; it must supply (i) a voltage equal to the D.P. across the resistance, $\mathcal{V}_R = \mathcal{I}R$ (ii) a voltage equal to the counter-E.M.F. of the inductance, $\mathcal{V}_L = \omega L\mathcal{I}$ and (iii) a voltage equal to the counter-E.M.F. of the condenser, $\mathcal{V}_c = \frac{\mathcal{I}}{\omega C}$. The three components of the

total E.M.F. $\mathcal{E}$, are shown in the vector diagram, from which it will be seen that the vector $\mathcal{V}_R$ is in phase with the current, the vector $\mathcal{V}_L$ is 90° leading, and the vector $\mathcal{V}_C$ 90° lagging on the vector $\mathcal{I}$. This must be interpreted as signifying that the vectors $\mathcal{V}_C$ and $\mathcal{V}_L$ partly cancel each other, the P.D. of the condenser assisting, during certain portions of each cycle, to create or destroy the magnetic field round the inductance, while in turn the E.M.F. set up by the changing magnetic field around the inductance assists in charging the condenser. The out-of-phase or reactive component of the E.M.F. has only to supply the difference between the vectors $\mathcal{V}_L$ and $\mathcal{V}_C$ and will be equal to $\omega L \mathcal{I} - \frac{\mathcal{I}}{\omega C}$. It may be noted that the second term of this voltage may in

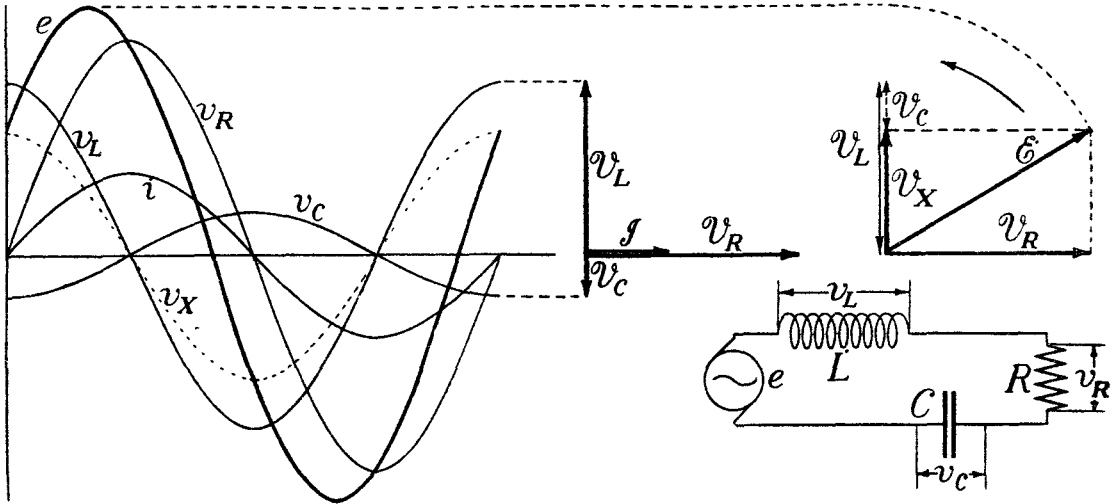


FIG. 25, CHAP. V.—Effect of resistance, inductance and capacitance in series.

some instances be larger than the first. If $\omega L \mathcal{I}$ is greater than $\frac{\mathcal{I}}{\omega C}$ the reactive voltage will be positive and will lead on the current, while if $\frac{\mathcal{I}}{\omega C}$ is greater than $\omega L \mathcal{I}$ the reactive voltage will be negative and will lag on the current. The effective reactive voltage being denoted by $\mathcal{V}_X$, the vector diagram shows that $\mathcal{E} = \sqrt{\mathcal{V}_R^2 + \mathcal{V}_X^2}$ that is, $\mathcal{E} = \mathcal{I} \sqrt{R^2 + \left(\omega L - \frac{1}{\omega C}\right)^2}$ or in R.M.S. values

$$I = \frac{E}{\sqrt{R^2 + \left(\omega L - \frac{1}{\omega C}\right)^2}}$$

the quantity $\sqrt{R^2 + \left(\omega L - \frac{1}{\omega C}\right)^2}$ is called the total impedance of the circuit.

Example 6.—A condenser of .0015 microfarads, a resistance of 20 ohms and an inductance of 25 microhenries are connected in series, and an alternating E.M.F. of 1.5 millivolts R.M.S., at a frequency of 10^6 cycles per second, is applied to the circuit. Find (i) the R.M.S. value of the current (ii) the R.M.S. voltages across the condenser and inductance respectively, and (iii) the angle of phase difference.

-CHAPTER V.—PARA. 27

It will be noted that the circuit is the same as in example 5 except that an inductance has been inserted in series.

$$\omega = 2\pi f = 2\pi \times 10^6$$

$$X_c = \frac{1}{\omega C} = 106 \text{ ohms}$$

$$X_L = \omega L = 2\pi \times 10^6 \times 25 \times 10^{-6} = 157 \text{ ohms}$$

$$X = \omega L - \frac{1}{\omega C} = 157 - 106 = 51 \text{ ohms}$$

$$Z = \sqrt{R^2 + X^2}$$

$$= \sqrt{20^2 + 51^2}$$

$$= \sqrt{3001}$$

$$= 54.8 \text{ ohms.}$$

$$(i) \quad I = \frac{E}{Z} = \frac{1.5}{54.8} \text{ milliamperes}$$

$$= .0274 \text{ milliamperes.}$$

In example 5 the R.M.S. current was only .0139 milliampere. The introduction of the inductance has caused an increase of current amounting to about 100 per cent.

$$(ii) \quad V_c = IX_c = \frac{I}{\omega C}$$

$$= .0274 \times 106 \text{ millivolts}$$

$$= 2.9 \text{ millivolts}$$

$$V_L = IX_L = \omega LI$$

$$= .0274 \times 157 \text{ millivolts.}$$

$$= 4.3 \text{ millivolts.}$$

Both V_L and V_c are greater than the R.M.S. applied voltage. The total reactive voltage is $V_L - V_c$ or 1.4 millivolts R.M.S., and this is equal to the reactive component of the alternator E.M.F.

(iii) The total reactance is $X_L - X_c = 51$ ohms. Since X_L is greater than X_c , the current will lag on the applied E.M.F. by an angle $\tan^{-1} \frac{X}{R}$.

$$\frac{X}{R} = \frac{51}{20} = 2.55$$

$$\theta = \tan^{-1} 2.55 = 69^\circ \text{ nearly.}$$

Effective inductance or capacitance

27. From the foregoing it will be observed that the reactive voltage $I \left(\omega L - \frac{1}{\omega C} \right)$ may be either leading or lagging on the current. Now inductive reactance causes the voltage to lead, and capacitive reactance causes it to lag. If $\omega L - \frac{1}{\omega C}$ is positive, it must have the same effect as an inductive reactance ωL_e , where L_e is the effective inductance of the components L and C in series, and

$$\omega L - \frac{1}{\omega C} = \omega L_e$$

$$\text{or } L_e = L - \frac{1}{\omega^2 C}$$

The effect of the capacitance is, in this particular instance, to reduce the apparent inductance L by an amount $\frac{1}{\omega^2 C}$. This is apparent from fig. 25, for the angle of phase difference is less with the condenser in circuit than if it were absent, and the impedance is also decreased with a proportional increase of current. If however the reactance $\omega L - \frac{1}{\omega C}$ is negative, the total effect of L and C must be that of a condenser, the equivalent value of which can now be found. If the latter is denoted by C_e ,

$$\begin{aligned}\omega L - \frac{1}{\omega C} &= - \frac{1}{\omega C_e} \\ \omega^2 L C - 1 &= - \frac{C}{C_e} \\ \text{or } C_e &= \frac{C}{1 - \omega^2 L C}\end{aligned}$$

Example 7.—(i) In example 5 what is the effective inductance of the circuit ?
The effective reactance has been found to be 51 ohms.

$$\begin{aligned}\omega L_e &= 51 \\ \omega &= 2\pi \times 10^6 \\ \therefore L_e &= \frac{51}{2\pi \times 10^6} \text{ henries} \\ &= \frac{51}{2\pi} \text{ or } 8.1 \text{ microhenries.}\end{aligned}$$

(ii) If in example 6 the frequency is changed to 0.6×10^6 what is the effective capacitance of the circuit ?

Instead of proceeding as above, the formula $C_e = \frac{C}{1 - \omega^2 L C}$ may be used.

$$\begin{aligned}\omega &= 2\pi \times 0.6 \times 10^6 \\ &= 3.77 \times 10^6 \\ \omega L &= 3.77 \times 10^6 \times 25 \times 10^{-6} \\ &= 94.25 \\ \omega C &= 3.77 \times 10^6 \times 0.0015 \times 10^{-6} \\ &= 0.00565 \\ \omega^2 L C &= 94.25 \times 0.00565 \\ &= 0.532 \\ 1 - \omega^2 L C &= 0.468 \\ \therefore C_e &= \frac{C}{0.468} \\ &= \frac{0.0015}{0.468} \text{ microfarads} \\ &= 0.0032 \text{ microfarads.}\end{aligned}$$

It has now been shown that a circuit possessing both capacitance and inductance in addition to its resistance, may behave at certain frequencies as though it possessed no capacitance, but an amount of inductance smaller than that actually existent, while at other frequencies it may behave as though it possessed no inductance, but an amount of capacitance greater than the actual value.

POWER MEASUREMENT IN HEAVY-CURRENT PRACTICE

28. The amount of power which is dissipated in a direct current circuit is given by the equation $P = I^2 R$. In an A.C. circuit the power is given by an identical expression, provided that the R.M.S. value of the current is employed, because this value is by definition the direct current which is equivalent in heating effect to the given alternating current. In an A.C. circuit possessing resistance only, the power may also be calculated from the product of R.M.S. amperes and R.M.S. volts, because the R.M.S. voltage is defined by means similar to those adopted for the definition of R.M.S. current. In circuits possessing reactance, however, the product of volts and amperes does not give the true power expended, but a quantity called the activity, apparent power, or simply the volt-amperes. In a circuit possessing both capacitive and inductive reactance, as well as resistance, it has been stated that the applied E.M.F. consists of three components, namely (i) $v_R = i R$ which is required to overcome the resistance, and is in phase with the E.M.F. (ii) $v_L = \omega L i$ which overcomes the counter-E.M.F. of self-induction. This component may be considered to establish the magnetic field in and around the coils constituting the inductance. (iii) $v_C = \frac{i}{\omega C}$ which is devoted to the establishment of an electric field between the plates of the condenser.

No average power is supplied from the source of E.M.F. in order to maintain the magnetic and electric fields; the energy required to establish them is returned to the source on their destruction. The voltages v_L and v_C are 90° out of phase with the current, and are referred to as wattless components. The R.M.S. current is given by the equation $I = \frac{E}{Z}$ and the power expended in the circuit by $I^2 R$. Hence $P = \frac{E^2}{Z^2} R = E \frac{E R}{Z^2} = E I \cos \phi$, where ϕ is the phase difference and may be either a leading or lagging angle.

In any A.C. circuit a hot-wire or electrostatic voltmeter connected across the supply terminals will give the R.M.S. value of the terminal P.D., V , and a hot-wire ammeter in series with the consuming device will read the R.M.S. current, I . The product of these readings gives the apparent power, $V I$, and the true power is the product multiplied by $\cos \phi$. The numeric $\cos \phi = \frac{R}{Z}$ is therefore called the Power Factor of the circuit. True power may however be measured directly by means of a wattmeter, and the Power Factor may be determined by the relation

$$\begin{aligned} \cos \phi &= \frac{\text{True power}}{\text{apparent power}} \\ &= \frac{\text{wattmeter reading}}{\text{Product of voltmeter and ammeter readings.}} \end{aligned}$$

The wattmeter

29. Two principal types of wattmeter are in use, and are known as the dynamometer and induction types respectively. Hot-wire and electrostatic types have also been proposed but have not been developed into practical instruments. The dynamometer instrument is similar in principle to the moving coil D.C. instrument, but contains no permanent magnet. It is shown diagrammatically in fig. 26. An alternating magnetic field is set up by a fixed coil carrying the main current or a definite fraction thereof, and the strength of this field, at any instant, is proportional to the instantaneous current. The moving coil is situated in this field, and is connected across the supply mains with a suitable resistance in series. The current in this coil is therefore proportional to the terminal P.D. Thus the connections of the fixed coil resemble those of an ammeter and the connections of the moving coil those of a voltmeter. No iron is used in the vicinity of the coils.

The torque exerted upon the moving coil is proportional to the product of the two currents and therefore proportional to the product of the main current and terminal P.D., that is to the

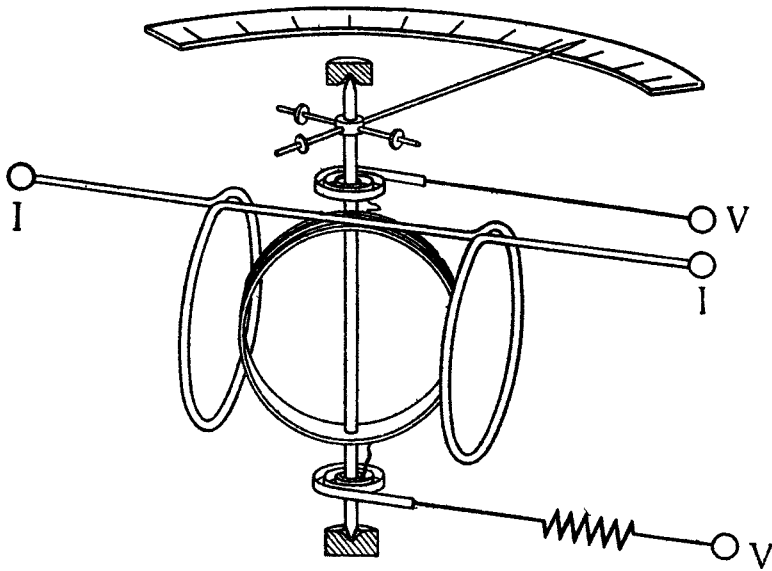


FIG. 26, CHAP. V.—Principle of dynamometer instrument.

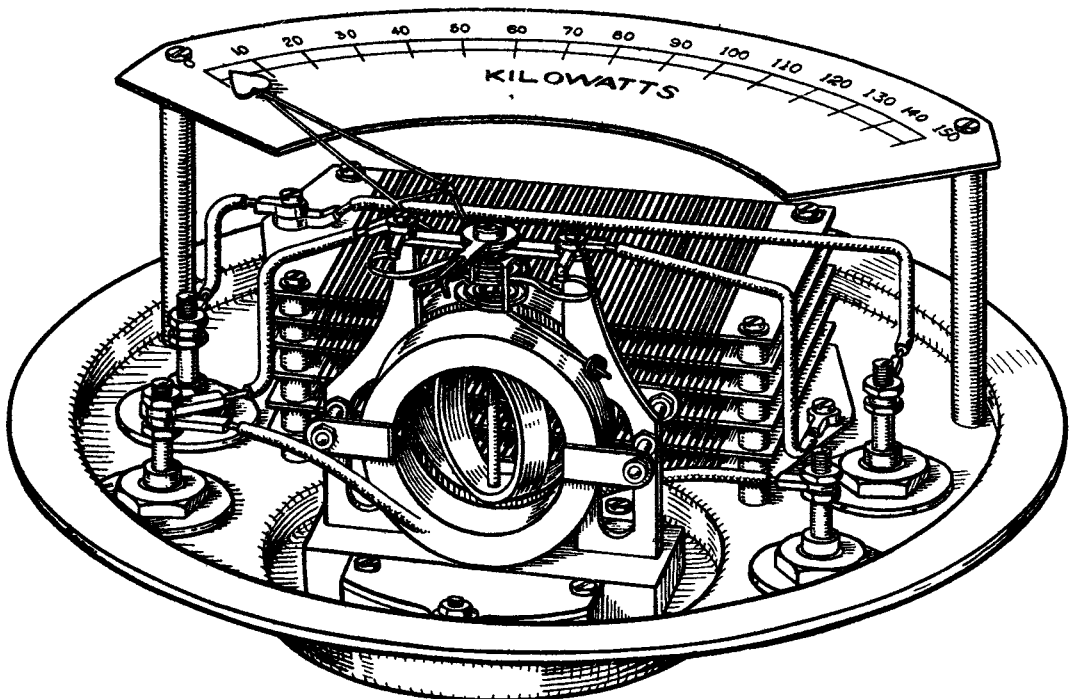


FIG. 27, Chap. V.—Dynamometer wattmeter.

CHAPTER V.—PARAS. 30-31

instantaneous power. The moving system cannot follow the rapid changes in the latter quantity, however, but takes up some position in which the controlling torque is equal to the mean value of the deflecting torque, that is the mean power, VI .

The controlling torque is obtained by spiral springs which also serve as connecting leads to the moving coil, and the movement is made deadbeat by an air damping device. The general appearance of a typical wattmeter of this type is shown in fig. 27.

The induction wattmeter

30. In this type of instrument the moving member consists of a thin aluminium disc, which is mounted upon a spindle, and is free to rotate through about 300° against the action of a light spiral spring. An electromagnet is arranged on each side of the disc, the poles of each being opposite to a different portion. The winding of one electromagnet carries the main current or a known fraction thereof, while the other winding is connected across the mains and carries a current which is proportional to the P.D. The latter winding has a series inductive resistance which causes its current to lag by nearly 90° upon the terminal P.D. The alternating fluxes set

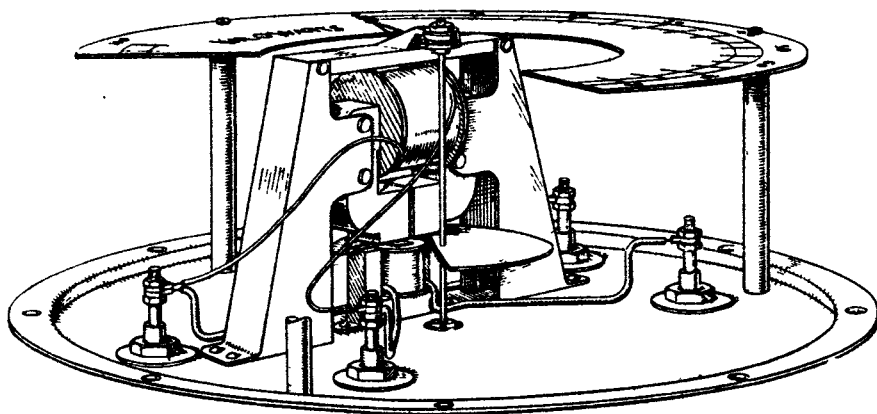


FIG. 28, CHAP. V.—Induction wattmeter.

up by these two magnets induce local electromotive forces in the disc, and consequently eddy currents are set up in the latter. These eddy currents in turn set up a flux which reacts upon the original flux, and a torque is exerted upon the disc which is proportional to the product of the fluxes caused by the two electromagnets and also to the sine of the angle of phase difference between them. Thus the torque is proportional to $VI \sin(90^\circ - \phi)$ or $VI \cos \phi$. A diagrammatic representation of this instrument is given in fig. 28. The instrument requires certain compensating devices (not shown on this figure) in order to avoid inaccuracy in indication, because it is impossible to make the parallel winding so highly inductive (or slightly resistive) that the current in it will lag by 90° on the terminal P.D.

Power measurement by voltmeter

31. When it is necessary to measure power or power factor in an A.C. circuit and a wattmeter is not fitted, the same information can be obtained in the following manner.

In fig. 29 (S_1) (S_2) are the supply terminals, and Z is the device whose power or power factor are to be measured. In series with Z is connected a noninductive resistance R such as a number of carbon filament lamps or an electric radiator; a hot-wire voltmeter (including of course its series resistance) is arranged to read at will either V_1 , the P.D. between the terminals of the device Z ; V_2 , the P.D. between the terminal of the noninductive resistance R , or V_3 the P.D. between the terminals of R and Z in series. An ammeter is connected in the supply line.

The relation between the voltages V_1 , V_2 and V_3 can be shewn by a vector diagram, which has been drawn beside the circuit diagram for easy reference. The vector I represents the current; the P.D. V_2 across the resistance is in phase with I , while the voltage V_1 across Z leads upon I by some angle ϕ . The supply voltage V_3 is the vector sum of V_1 and V_2 . From the end of the vector V_3 a perpendicular to the current vector is drawn, this being shown as a dotted line in the figure. This completes a right-angled triangle the hypotenuse of which is V_3 . By inspection we find that the two other sides of this triangle are $(V_1 \cos \phi + V_2)$ and $(V_1 \sin \phi)$. From this information we deduce that

$$V_3^2 = (V_1 \cos \phi + V_2)^2 + (V_1 \sin \phi)^2$$

$$V_3^2 = V_1^2 \cos^2 \phi + 2 V_1 V_2 \cos \phi + V_2^2 + V_1^2 \sin^2 \phi$$

But

$$\sin^2 \phi + \cos^2 \phi = 1$$

Hence

$$V_3^2 = V_1^2 + V_2^2 + 2 V_1 V_2 \cos \phi$$

and

$$V_1 \cos \phi = \frac{V_3^2 - V_1^2 - V_2^2}{2 V_2}$$

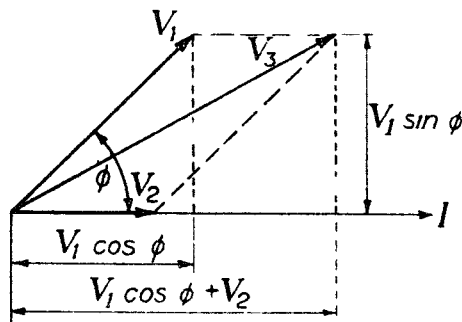
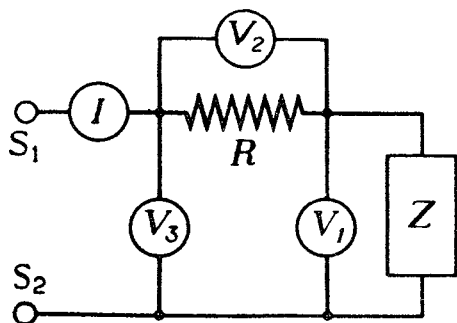


FIG. 29, CHAP. V.—Power measurement in A.C. circuit (Three voltmeter method).

The current flowing, as shown by the ammeter, can now be introduced giving

$$V_1 I \cos \phi = \frac{V_3^2 - V_1^2 - V_2^2}{2 V_2} I$$

which is the power consumed by the device Z . As the power is deduced from the differences between the squares of quantities, small errors in these quantities, i.e. in the voltmeter readings, produce considerably greater errors in the value of the power. For best results the series resistance should be so chosen that its value is about equal to the impedance of the device under measurement. This necessitates a supply voltage at least 50 per cent. higher than that for which Z is rated, while in the unlikely event of the latter being completely non-reactive, the normal supply voltage would have to be doubled for the purposes of the measurement. This disadvantage can be overcome by using the following method of measurement.

Power measurement by ammeter

32. The general principles of this method are similar to those in the preceding. The connections are shown in fig. 30 in which the device under measurement is again Z . Ammeters are connected in each branch of the circuit, the noninductive resistance R with its meter being placed in parallel with Z . From the vector diagram it is deduced in exactly the same manner as before that

$$V I_1 \cos \phi = \frac{I_3^2 - I_2^2 - I_1^2}{2 I_2} V$$

CHAPTER V.—PARAS. 33-34

The disadvantage of this method is that three ammeters of identical accuracy are required, otherwise elaborate switching arrangements are necessary in order to transfer the ammeter from circuit to circuit. Both methods assume that the wave form is sinusoidal and inaccurate results are obtained if this assumption is incorrect.

33. Power factor meters may be used to indicate directly the power factor of a circuit. A typical design consists of a fixed coil which carries the main current, this coil being divided into two halves. Pivoted in the space between these is the moving element consisting of two coils rigidly fixed at right-angles to each other so that they move as one unit. No controlling torque is required. One coil of the moving element in series with a noninductive resistance is connected across the supply mains, the winding of the other being similarly connected but with a reactive device, i.e. a condenser or inductance, in series, and consequently the former winding carries a current which is proportional to and in phase with the voltage, while the latter winding carries a current which is proportional to the voltage but which differs considerably in phase. It will be considered as a lagging current in the following explanation.

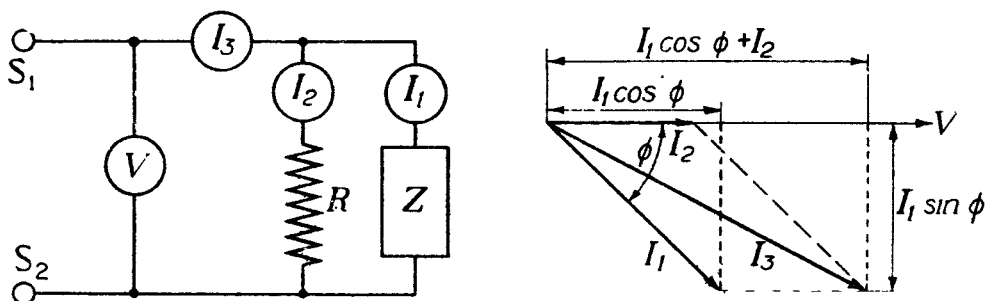


FIG. 30, CHAP. V.—Power measurement in A.C. circuit (Three ammeter method).

Suppose the main circuit to have unity power factor, then the current in the non-inductive branch of the moving unit will interact with the main field in such a way that its coil will set itself parallel to the main coil so that it embraces maximum flux; on the other hand if the lag of current in the main coil is exactly equal to the lag of current in the inductive branch of the moving unit, the latter will be turned until the coil embraces maximum flux, and will be parallel to the main coil. For any intermediate angles of lag the moving unit takes up an intermediate position in which the resultant field of the moving unit is parallel to the main field at the instant of peak value of the latter. With a leading current in the main coil, the same arguments apply, except that the moving coil in the inductive branch will turn in the opposite direction to that in which it turned with a lagging main current. The pointer attached to the moving element moves over a scale graduated up to 90° on either side of zero, thus showing the phase angle and whether the current is leading or lagging.

Energy meters

34. Energy meters for A.C. supply are usually true watt-hour meters. The typical form is similar in principle to the induction type of wattmeter. The chief modification is the removal of the spiral spring constituting the controlling force of the wattmeter, so that the spindle carrying the disc is free to rotate in its bearings, and the provision of a counting mechanism, e.g. a cyclometer, instead of a pointer. This cyclometer is driven from the spindle of the disc by a worm cut in the latter.

A retarding torque is provided by an additional permanent magnet, by which eddy currents are induced in the revolving disc, and consequently this torque is directly proportional to the speed, while the torque exerted by the electromagnets is proportional to the power supplied or to $VI \cos \phi$. Hence the total number of revolutions is proportional to the watt-hours.

Rating of alternating current machinery

35. Makers of alternating current machinery rate their products as being capable of delivering a given number of kilo-volt-amperes instead of a given number of kilo-watts, e.g. an alternator may be spoken of as a 200 volt, 10 kVA machine. This means that at its rated speed it will deliver the rated voltage and is capable of delivering 10,000 volt-amperes without overheating. If the load is purely inductive, this current will be wattless, and no power will be dissipated in the external circuit, although the machine is giving an output of 10 kVA and the internal losses are exactly the same as when current is delivered to a power-dissipating circuit. On the other hand if the maker guaranteed his alternator to produce 10 kW at 200 volts irrespective of the nature of the load, and the machine were called upon to deliver this power to a load having a power factor of 0.5, the apparent power would be $\frac{10,000}{0.5} = 20,000$ volt-amperes, and the current would be $\frac{20,000}{200} = 100$ amperes, or double the current required by a load possessing a power factor of unity. With the reactive load the heating effect in the machine itself is obviously four times that caused by the non-reactive load, and the machine would certainly suffer damage.

SERIES RESONANCE

36. We have seen that under certain conditions the inductive and capacitive effects in a circuit tend to cancel each other. This cancellation is complete if the values of L , C and ω are such that $\omega L - \frac{1}{\omega C} = 0$. The counter-E.M.F. of self-induction, ωLi , and the P.D. at the condenser terminals, $\frac{i}{\omega C}$, are then equal in magnitude and opposite in phase at every instant throughout the cycle, and the circuit will behave as if it had neither inductance nor capacitance. The circuit is then said to be in series resonance with the frequency of the applied E.M.F.

The resonant frequency of a circuit possessing capacitance and inductance in series may be defined as the frequency at which the total reactance is zero. For given values of L and C the resonant frequency f_r is found by equating ωL to $\frac{1}{\omega C}$ thus:—

$$\begin{aligned}\omega L &= \frac{1}{\omega C} \\ \omega^2 &= \frac{1}{LC} \\ \omega &= \frac{1}{\sqrt{LC}} = 2\pi f_r \therefore f_r = \frac{1}{2\pi\sqrt{LC}}\end{aligned}$$

As the reactance of a circuit to an E.M.F. at its resonant frequency is zero, the current at this frequency must depend only upon the resistance of the circuit, and is given by the equation

$$i = \frac{E \sin \omega t}{R}$$

the R.M.S. value being

$$I = \frac{E}{R}$$

This gives an alternative definition of series resonance, i.e. the frequency at which the current has the value $\frac{E}{R}$. In a resonant circuit, the current and E.M.F. are in phase, and the power factor is unity.

The term resonance is borrowed from the science of acoustics, and numerous examples occur in all branches of physics, many being matters of everyday experience. For instance, a wind

CHAPTER V.—PARA. 37

glass which emits a clear note of definite pitch when tapped sharply, will emit a similar sound if the appropriate note is produced by a musical instrument in the vicinity. The action of the reed type frequency meter (para. 7) depends upon this principle, the natural frequency of vibration of one particular reed being coincident with the frequency of the current flowing through the magnet winding.

Electrical resonance is not often met with in heavy current engineering, and when a possibility of resonance exists it is generally suppressed by suitable variation of the circuit constants. It will be shown that under resonant conditions the voltage across certain circuit components may be many times the applied voltage, and in power circuits this increase of voltage would necessitate very heavy insulation, besides leading to other complications into which it is unnecessary to enter. In radio circuits, however, and particularly in receivers where the applied voltage is often only of the order of a few microvolts, the phenomenon of resonance is utilised in order to achieve effects greater than could be obtained by direct employment of the available voltage. For this reason further discussion of resonance will be illustrated by examples of direct application to radio-communication. The terms "audio-frequency" and "radio-frequency" have already been introduced, the range included in the latter term being from 20,000 cycles per second to several million cycles per second. It is convenient to refer to frequencies of this order in kilocycles per second (kc/s) or megacycles per second (Mc/s), while it has also been proposed to use the term hertz to denote one cycle per second, but this unit has not yet been adopted for service use. In radio-frequency circuits the inductance and capacitance are invariably only small fractions of a henry and farad respectively, and the units used are the microhenry and microfarad.

It has been shown that the resonant frequency of a circuit possessing an inductance of L henries and a capacitance of C farads is given by the formula

$$f_r = \frac{1}{2\pi \sqrt{LC}}$$

It is often more convenient to use a formula giving the resonant frequency in terms of the inductance in microhenries, and the capacitance in microfarads. This is derived as follows:—

Let L = the inductance of the circuit, in microhenries.

C = the capacitance of the circuit, in microfarads.

f_r = the resonant frequency of the circuit.

Since 1 henry = 10^6 microhenries

1 farad = 10^6 microfarads

$$\begin{aligned} f_r &= \frac{1}{2\pi \sqrt{\frac{L}{10^6} \times \frac{C}{10^6}}} \\ &= \frac{10^6}{2\pi \sqrt{LC}} \end{aligned}$$

When this relation is satisfied, the series circuit is said to be an acceptor circuit for the frequency f .

Series resonance curves

37. Referring to the circuit shown in fig. 25, let $L = 150 \mu H$, $C = 000169 \mu F$, $R = 10$ ohms and the E.M.F. of the alternator to be 10 millivolts (R.M.S.). Suppose the frequency to be variable between say 950 kc/s and 1,050 kc/s. The current at any frequency

$f = \frac{\omega}{2\pi}$ is given by the equation

$$I = \frac{E}{\sqrt{R^2 + \left(\omega L - \frac{1}{\omega C}\right)^2}}$$

bearing in mind that L denotes the inductance in henries and C the capacitance in farads. As the frequency is varied between the given limits, the current will also vary; its value has been calculated over the range 970 to 1,030 kc/s, and the results plotted in fig. 31 curve (i). It will be seen that the current reaches the value $\frac{E}{R}$, i.e. one milliampere, when the supply frequency is 1,000 kc/s, which is the resonant frequency. On either side of resonance, the current is less than this, falling off rapidly at first and then more slowly. At frequencies below 1,000 kc/s the, capacitive reactance $\frac{1}{\omega C}$ is greater than the inductive reactance ωL and the current leads upon

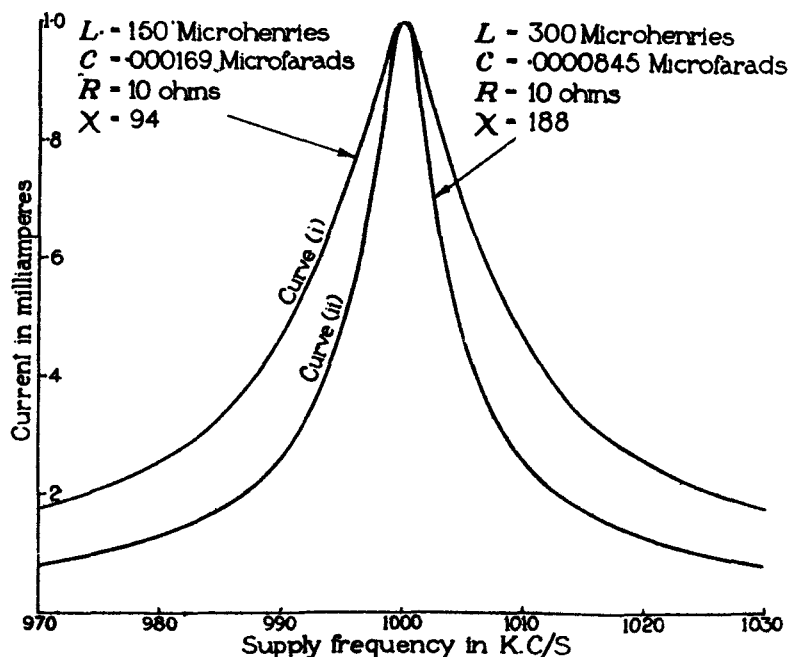


FIG. 31, CHAP. V.—Series resonance curves. Effect of ratio of inductance to capacitance.

the applied voltage, while at frequencies above 1,000 kc/s the inductive reactance is greater than the capacitive reactance and the current lags on the applied voltage. The graph showing the variation of current as the frequency is varied is called the resonance curve of the circuit.

Selectivity of an acceptor circuit

38. One of the principal applications of the phenomenon of electrical resonance is in the radio receiving circuit. A distant radio transmitter sets up in a receiving aerial an alternating E.M.F., the frequency of which is the same as that of the transmitter. More complete consideration of both transmitters and receivers will be found in subsequent chapters, but for the present it may be considered that all transmitters of equal power, situated at the same distance from the receiver, may be expected to produce equal E.M.F.'s in the receiving aerial. (Certain qualifications of this assumption are necessary and will be found in the appropriate chapters.) The receiving aerial circuit possesses inductance, capacitance, and resistance and may be represented as in fig. 32, where the alternators E_1 , E_2 , E_3 represent induced E.M.F.'s, having frequencies f_1 , f_2 , f_3 respectively. These alternators therefore give the same effect in the circuit as three different transmitters, and if $f_1 = 990$ kc/s, $f_2 = 1,000$ kc/s and $f_3 = 1,020$ kc/s, fig. 31 curve (i)

CHAPTER V.—PARA. 39

shews that if E_1 , E_2 and E_3 are each equal to 10 millivolts, E_1 will produce a current of $\cdot 46$ milliampere, E_2 a current of 1 milliampere and E_3 a current of $\cdot 26$ milliampere. Although the three voltages applied to the circuit are of equal value, the one which has a frequency equal to the resonant frequency of the circuit will produce the largest current, and therefore the strongest signal in the telephone receivers or loud speaker of the receiver. The inclusion of an acceptor

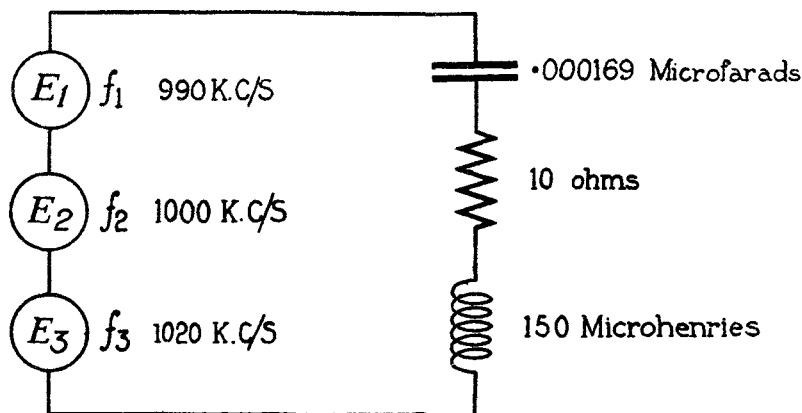


FIG. 32, CHAP. V.—Series L/C circuit with applied E.M.F. at resonant and non-resonant frequencies.

circuit in a piece of apparatus thus endows it with the property of discrimination in favour of signals (i.e. E.M.F.'s) of resonant frequency, to the partial exclusion of others. This capability to differentiate between signals of different frequencies is called selectivity, and the selectivity of an acceptor circuit depends upon its ratio of inductance to resistance.

Influence of ratio $\frac{L}{R}$

39. The effect of an increase in the ratio $\frac{L}{R}$ is also shown in fig. 31. Curve (i) has already been referred to; it is the resonance curve for a ratio $\frac{L}{R}$ of $\frac{150}{10} = 15$. In curve (ii) the inductance has been increased to $300 \mu H$, the capacitance being correspondingly reduced so that the resonant frequency remains 1,000 kc/s, the ratio $\frac{L}{R}$ being 30. The three E.M.F.'s cited above would now produce the following currents, viz: at 990 kc/s, $\cdot 23$ milliampere; at 1,000 kc/s 1 milliampere, as before, and at 1,020 kc/s $\cdot 13$ milliampere. With this ratio of $\frac{L}{R}$, then, the current at resonance is unchanged, but E.M.F.'s of non-resonant frequencies cause considerably reduced currents to flow, and the selectivity is therefore increased.

The ratio $\frac{L}{R}$ may also be changed by a variation of resistance, instead of by variation of inductance. Fig. 33 shows resonance curves for fixed values of inductance and capacitance ($150 \mu H$ and $\cdot 000169 \mu F$ respectively); curve (i) representing the state of affairs when the resistance is 10 ohms, is repeated from fig. 31 for comparison. Curve (ii) shows the effect of increasing the resistance to 14.14 ohms and curve (iii) the effect of decreasing the resistance to 7.07 ohms. Comparing the two latter curves, it is seen that at the resonant frequency, halving the resistance results in doubling the current, but at any other frequency the effect is not so

marked. At 970 kc/s and at 1,030 kc/s the current is practically independent of the resistance and depends only upon the reactance of the circuit, so that the current is the same in all three cases.

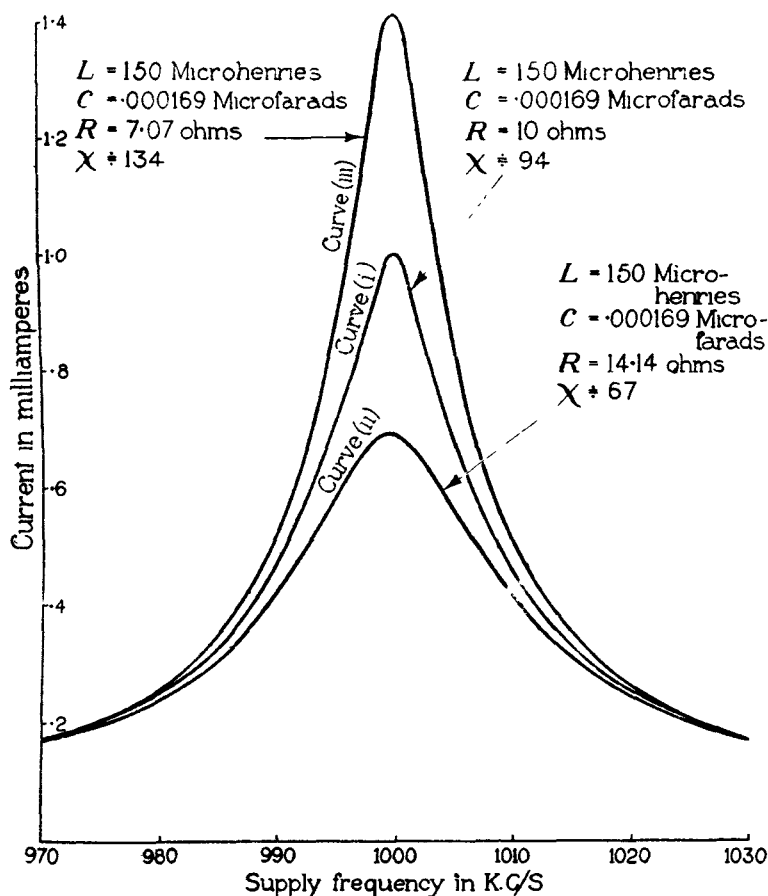


FIG. 33, CHAP. V.—Series resonance curves. Effect of variation of resistance.

40. The reader must be on his guard against a common fallacy, i.e. that an increase in the ratio $\frac{L}{C}$ will necessarily lead to an increase in the selectivity of an acceptor circuit. This is not true unless the increase of inductance is accompanied by an increase in the ratio $\frac{L}{R}$ which in itself is sufficient to cause an increase of selectivity. As an example, consider the selectivity of two different acceptor circuits. Let R_1, L_1, C_1 be the constants of one circuit, and R_2, L_2, C_2 those of the other, also let $R_2 = 2R_1, L_2 = 2L_1, C_2 = \frac{1}{2}C_1$; the circuits then have the same resonant frequency and the same ratio of inductance to resistance, but the ratio of inductance to capacitance in the circuit R_2, L_2, C_2 is four times as great as in the circuit R_1, L_1, C_1 . Let I_r denote the current at the resonant frequency $\frac{\omega_r}{2\pi}$, and I_n the current at any other frequency, $\frac{\omega_n}{2\pi}$.

CHAPTER V.—PARA. 41

In the first circuit

$$I_r = \frac{E}{R_1}, I_n = \frac{E}{\sqrt{R_1^2 + X_1^2}} = \frac{E}{Z_1},$$

$$\text{where } X_1 = \omega_n L_1 - \frac{1}{\omega_n C_1}.$$

In the second circuit

$$I_r = \frac{E}{R_2}, I_n = \frac{E}{\sqrt{R_2^2 + X_2^2}}; X_2 = \omega_n L_2 - \frac{1}{\omega_n C_2}.$$

$$X_2 = 2\omega_n L_1 - \frac{2}{\omega_n C_2} = 2X_1; \therefore I_n = \frac{E}{\sqrt{(2R_1)^2 + (2X_1)^2}} = \frac{E}{2Z_1}.$$

For the present purpose the selectivity may be defined as the ratio of current at resonance to current at the non-resonant frequency, that is $\frac{I_r}{I_n}$.

In the first circuit

$$\frac{I_r}{I_n} = \frac{E}{R_1} \div \frac{E}{Z_1} = \frac{\sqrt{R_1^2 + X_1^2}}{R_1}.$$

In the second circuit

$$\frac{I_r}{I_n} = \frac{E}{2R_1} \div \frac{E}{2Z_1} = \frac{\sqrt{R_1^2 + X_1^2}}{R_1};$$

as I_n is the current at any non-resonant frequency whatever, it is seen that although the maximum value of the current is different, the two resonance curves will have exactly the same shape, and the two circuits have the same selectivity, because the ratio $\frac{I_r}{I_n}$ is the same in each example.

Voltage magnification

41. It has already been shown that in a circuit possessing both capacitance and inductance connected in series, the terminal P.D. of the coil or condenser may be greater than the applied E.M.F. This effect is most pronounced when the frequency of the applied E.M.F. is identical with the resonant frequency of the circuit. Let E be the applied E.M.F. (R.M.S.) L the value of the inductance in henries, C the capacitance of the condenser in farads, and R the resistance of the circuit in ohms. If V_L is the P.D. at the terminals of the coil, V_C the P.D. at the condenser terminals, and $\omega_r = 2\pi$ times the resonant frequency,

$$V_L = \omega_r L I$$

$$V_C = \frac{1}{\omega_r C} I$$

$$\text{Since at resonance, } I = \frac{E}{R} \text{ and } \omega_r L = \frac{1}{\omega_r C},$$

$$V_L = V_C.$$

$$\begin{aligned} V_L^2 &= V_L V_C = \frac{\omega_r L E}{R} \times \frac{E}{\omega_r C R} \\ &= \frac{E^2 L}{R^2 C} \end{aligned}$$

$$\therefore V_L = V_C = \frac{E}{R} \sqrt{\frac{L}{C}}$$

Example 8.—A coil having an inductance of 160 microhenries, a condenser having a capacitance of .00025 microfarads, and a resistance of 10 ohms are placed in series with an E.M.F. of 2 volts (R.M.S.) at the resonant frequency. Find the R.M.S. voltage at the terminals of the coil and condenser.

$$\begin{aligned} V_L = V_C &= \frac{E}{R} \sqrt{\frac{L}{C}} \\ &= \frac{2}{10} \sqrt{\frac{160 \times 10^{-6}}{.00025 \times 10^{-6}}} \\ &= .2 \sqrt{640000} \\ &= 160 \text{ volts.} \end{aligned}$$

The ratio $\frac{V_L}{E}$ is called the resonance voltage magnification of the circuit, or if there is no danger of ambiguity, the circuit magnification. It may be denoted by the symbol χ . The appropriate values of this constant have been inserted on the curves in figs. 31 and 33. It will be seen that a circuit having a high value of χ gives a resonance curve which rises sharply as the resonant frequency is approached while with low values of χ the resonance curve tends to become flat. From this graphic point of view it has become the practice to speak of the sharpness of resonance; a circuit having high χ is said to be sharply resonant, and a circuit having low χ to be flatly resonant, or flatly tuned.

42. In practice the resistance of a radio-frequency circuit is generally an undesirable feature. In this respect wireless circuits differ from many alternating power circuits in which the desired effect is often the production of heat, sometimes for its own sake, as in electric radiators and soldering irons, and sometimes because the heat is required to render a body incandescent as in the electric lamp. The resistance of a radio circuit is often only that inherent in the coils and condensers, and the efficiency of the coil or condenser is given by the ratio of its reactance to its resistance. This ratio is often spoken of as the Q of the component, but as the symbol Q is used to denote quantity of electricity, the term figure of merit will be used to denote this ratio.

The figure of merit of an inductive coil is therefore equal to $\frac{\omega L}{R}$ where $\frac{\omega}{2\pi}$ is the frequency, L the inductance of the coil in henries and R its resistance in ohms. This figure is approximately constant over a wide frequency range owing to the fact that the h.f. resistance of a coil is roughly proportional to the frequency. Thus if R_1 is the resistance of the coil at a frequency $f_1 = \frac{\omega_1}{2\pi}$

and R_2 the resistance at a frequency $f_2 = \frac{\omega_2}{2\pi}$, $R_2 \doteq \frac{\omega_2}{\omega_1} R_1$. At the frequency f_1 the figure of merit is $\frac{\omega_1 L}{R_1}$ while at the frequency f_2 it is $\frac{\omega_2 L}{R_2}$ and $\frac{\omega_2 L}{R_2} \doteq \frac{\omega_2 L}{\frac{\omega_2}{\omega_1} R_1} \doteq \frac{\omega_1 L}{R_1}$.

The closeness of this approximation may be illustrated by the measured values for a certain coil, which possessed an inductance of 185 microhenries. At a frequency of 500 kc/s the figure of merit was 120, rising to 160 over the range 800 to 1,000 kc/s. At higher frequencies the figure of merit decreased slowly, being again 120 at 1,500 kc/s.

If an inductance is connected in series with a loss-free condenser, and an E.M.F. applied at the resonant frequency of the circuit, the resonant voltage magnification is equal to $\frac{\omega L}{R}$ i.e. to the figure of merit of the coil at this particular frequency. For this reason the term "coil magnification" is sometimes used instead of figure of merit, although this may lead to confusion. The symbol χ_L may be used to denote the figure of merit of a coil. The efficiency of a condenser may also be expressed as a figure of merit which is the ratio of its reactance to its effective resistance,

CHAPTER V.—PARAS. 43-44

including in the latter all sources of energy loss. It may be denoted by the symbol χ_c . As the energy losses are approximately in inverse proportion to the frequency, χ_c is again fairly constant over a wide frequency range.

PARALLEL COMBINATIONS OF INDUCTANCE, CAPACITANCE AND RESISTANCE

43. In drawing the vector diagrams to show the relative magnitudes and phase difference of current and voltage in a series circuit, the current vector is used as a datum line, because the same current flows through every component. The resistive P.D. R is drawn parallel to the vector $\mathcal{I}$ and the reactive P.D.'s perpendicular to the current vector, lagging or leading as the case may be. In parallel circuits the circuit components have a common terminal voltage and this vector is taken as the datum line; a vector representing the current through a resistance is drawn parallel to the voltage vector, and currents through purely reactive components perpendicular to the voltage vector, lagging or leading as requisite. The rule in drawing a vector diagram is therefore to use as a reference vector the one representing the quantity which is common to all circuit components. In the following paragraphs the discussion will be illustrated only with vector diagrams, although of course the corresponding sine and cosine curves could be used as in the case of series circuits.

Resistance and inductance in parallel

44. In fig. 34a is shown an inductance L and a resistance R connected in parallel, the inherent resistance of the inductance being negligible. An alternating E.M.F. of peak value $\mathcal{E}$ is

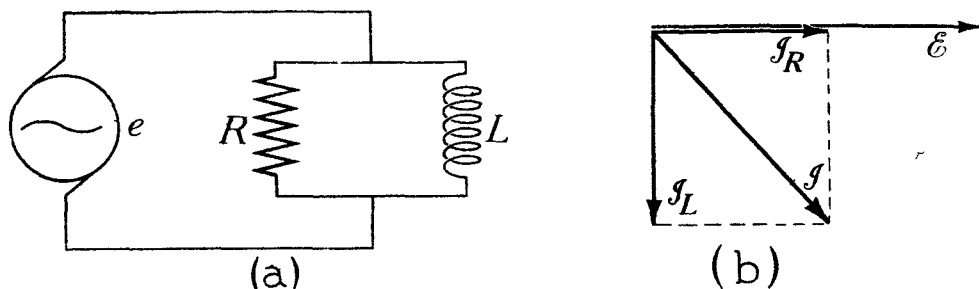


FIG. 34, CHAP. V.—Effect of resistance and inductance in parallel.

applied to the terminals of the combination, and consequently alternating currents of the same frequency will flow in both branches of the circuit. The current through the inductance will have a peak value $\mathcal{I}_L = \frac{\mathcal{E}}{\omega L}$ lagging on the applied voltage by 90° , while the current through the resistance will have the peak value $\mathcal{I}_R = \frac{\mathcal{E}}{R}$, and will be in phase with the applied voltage. The total current $\mathcal{I}$ supplied by the alternator will be the vector sum of $\mathcal{I}_L$ and $\mathcal{I}_R$, and is shown in the vector diagram fig. 34b, from which it is deduced that

$$\begin{aligned}\mathcal{I} &= \sqrt{\mathcal{I}_R^2 + \mathcal{I}_L^2} \\ &= \sqrt{\left(\frac{\mathcal{E}}{R}\right)^2 + \left(\frac{\mathcal{E}}{\omega L}\right)^2} \\ &= \mathcal{E} \sqrt{\left(\frac{1}{R}\right)^2 + \left(\frac{1}{\omega L}\right)^2}\end{aligned}$$

and this current will lag on the applied voltage by an angle

$$\theta = \tan^{-1} \frac{\mathcal{I}_L}{\mathcal{I}_R}, \text{ i.e.}$$

$$\tan \theta = \frac{\mathcal{I}_L}{\mathcal{I}_R} = \frac{\mathcal{E}}{\omega L} \times \frac{R}{\mathcal{E}} = \frac{R}{\omega L}.$$

In R.M.S. values,
$$I = E \sqrt{\left(\frac{1}{R}\right)^2 + \left(\frac{1}{\omega L}\right)^2}$$

and the ratio $\frac{E}{I}$ is the impedance of the parallel combination. Hence

$$Z = \frac{E}{I} = \frac{1}{\sqrt{\left(\frac{1}{R}\right)^2 + \left(\frac{1}{\omega L}\right)^2}}$$

The reciprocal of the impedance is called the admittance and is denoted by Y . The reciprocal of the resistance $\frac{1}{R}$, is termed the conductance and is denoted by G , while the reciprocal of the inductive reactance $\frac{1}{\omega L}$ is termed the inductive susceptance, and is denoted by B_L . (The symbol B is also used for flux density, but as flux density and susceptance rarely occur in the same calculation there is little risk of confusion.) The relation between E and I may therefore be written

$$I = YE = \sqrt{G^2 + B_L^2} E$$

Resistance and capacitance in parallel

45. For the inductance L in the preceding discussion let a capacitance C be substituted, and an E.M.F. of peak value $\mathcal{E}$ be applied to the parallel combination, fig. 35a. As before, an

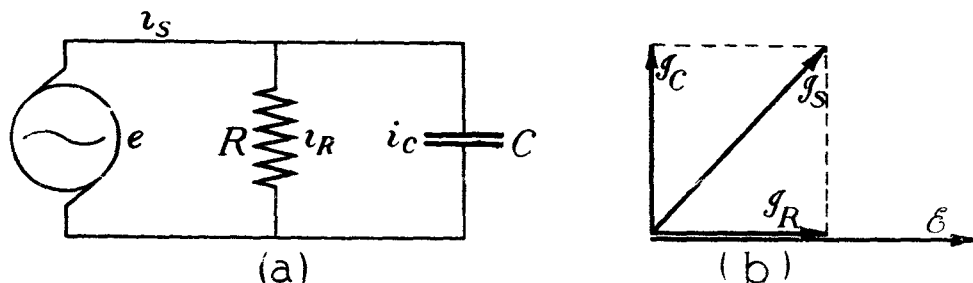


FIG. 35, CHAP. V.—Effect of resistance and capacitance in parallel.

alternating current of the supply frequency will flow in each component, the peak current through the resistance being $\mathcal{I}_R = \frac{\mathcal{E}}{R}$ in phase with the applied E.M.F. as before. The current charging the condenser will be $\mathcal{I}_C = \omega C \mathcal{E}$ and will lead on the applied E.M.F. by 90° . The total current $\mathcal{I}$ supplied by the alternator will be the vector sum of $\mathcal{I}_C$ and $\mathcal{I}_R$, and is shown in the vector diagram fig. 35b. It will be seen that

$$\begin{aligned} \mathcal{I} &= \sqrt{\left(\frac{\mathcal{E}}{R}\right)^2 + (\omega C \mathcal{E})^2} \\ &= \mathcal{E} \sqrt{\left(\frac{1}{R}\right)^2 + (\omega C)^2} \\ &= \mathcal{E} \sqrt{G^2 + B_C^2} \end{aligned}$$

CHAPTER V.—PARA. 46

and this current will lead on the applied voltage by an angle $\theta = \tan^{-1} \frac{g_c}{g_R} = \omega CR$. The expression

$\sqrt{\left(\frac{1}{R}\right)^2 + (\omega C)^2}$ is the admittance of the parallel combination, the capacitive susceptance being $B_C = \omega C$.

Example 9.—An inductance of 100 microhenries, having negligible resistance, and a resistance of 600 ohms, are placed in parallel. Find the R.M.S. current when an R.M.S. voltage of 2 volts is applied, at a frequency of 800 kc/s.

$$\omega = 2\pi f = 6.28 \times 800,000$$

$$\omega L = 6.28 \times 800,000 \times 100 \times 10^{-6} \text{ ohms}$$

$$R = 600 \text{ ohms}$$

$$I_L = \frac{E}{\omega L} = \frac{2}{502.4} = .00398 \text{ amperes or } 3.98 \text{ milliamperes.}$$

$$I_R = \frac{E}{R} = \frac{2}{600} = .00333 \text{ amperes or } 3.33 \text{ milliamperes.}$$

$$\begin{aligned} I &= \sqrt{I_R^2 + I_L^2} \\ &= \sqrt{3.33^2 + 3.98^2} \\ &= 5.2 \text{ milliamperes.} \end{aligned}$$

I_R is in phase with E , and I_L lags by 90° on E . Therefore I lags on E by an angle $\theta < 90^\circ$;

$$\tan \theta = \frac{I_L}{I_R} = \frac{3.98}{3.33} = 1.195$$

whence $\theta = 50^\circ$ approximately.

Or :— $G = \frac{1}{R} = .001667$

$$B_L = \frac{1}{\omega L} = .00199$$

$$\begin{aligned} Y &= \sqrt{G^2 + B_L^2} \\ &= \frac{\sqrt{2.77 + 3.95}}{1,000} = \frac{2.6}{1,000} \end{aligned}$$

$$I = YE = \frac{2.6}{1,000} \times 2 = .0052 \text{ amperes or } 5.2 \text{ milliamperes.}$$

$$\text{and } \tan \theta = \frac{B_L}{G} = \frac{600}{502} = 1.195$$

Impedance and resistance in parallel

46. The circuit of fig. 36 shows a resistance of 2 ohms, having an inductance of 0.4 microhenries, in parallel with a purely resistive path of 2 ohms. At very low frequencies, the

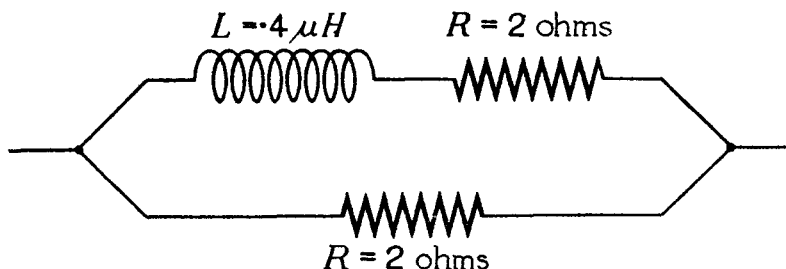


FIG. 36, CHAP. V.—Inductive resistance and non-inductive resistance in parallel.

inductive reactance is negligible and the current in each branch will be the same, for a given terminal P.D. Suppose however that the frequency is fairly high, say 796 kc/s. The inductive reactance is then also 2 ohms, and the impedance of this branch becomes $\sqrt{8}$ ohms, so that the current will be only $\cdot 707$ of that in the purely resistive branch. The disparity between the two currents will increase with the frequency, and it may be said that where parallel paths are available, a high frequency current will divide inversely as the inductance of the respective paths, the resistance having negligible effect upon the current distribution.

This leads to an alternative view of the cause of skin effect. The centre portion of the cross-section of a conductor is surrounded by a greater number of tubes of flux than the outer portion, and the peak value of the rate of change of flux increases from the surface to the centre, so that the inductance of the centre is greater than that of the outer portion. As the current divides in inverse ratio to the inductance, it follows that less current will flow in the centre of the cross-section than on the surface.

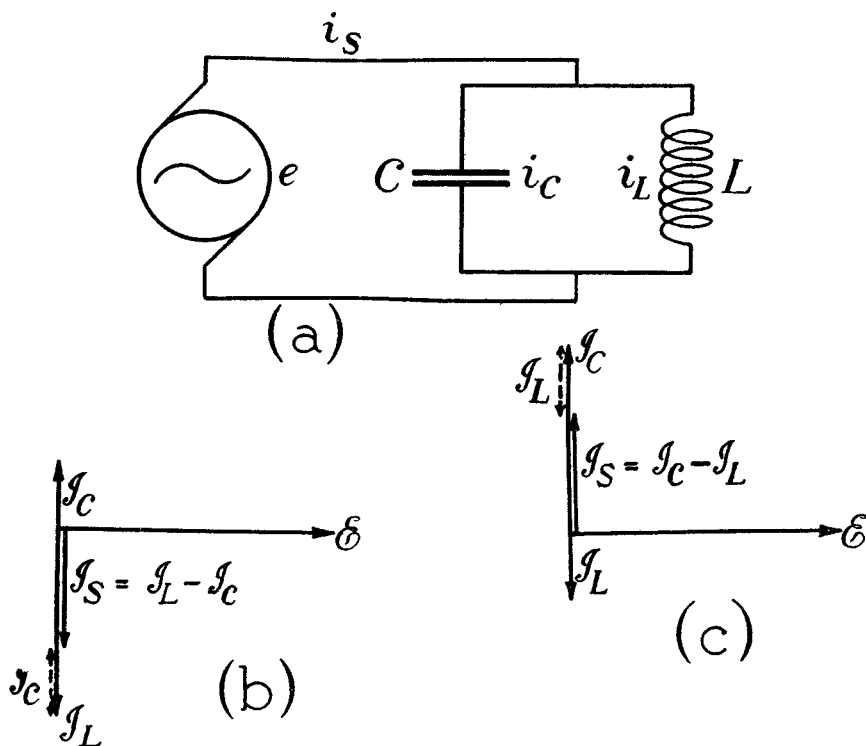


FIG. 37, CHAP. V.—A.C. circuit possessing inductance and capacitance in parallel.

Inductance and capacitance in parallel

47. It is next proposed to consider a circuit in which an inductance L and a condenser C are placed in parallel, and an alternating E.M.F. applied to the parallel combination. In the preliminary investigation, the effects of resistance will be neglected. It must not be supposed however that the results obtained in this way are valueless, for in practical circuits the resistance can often be reduced to a very small value. The circuit under consideration is shown in fig. 37a, in which the alternator E.M.F. has a peak value $\mathcal{E}$, while its frequency is variable. The peak value of the current flowing through the inductance will be $\mathcal{I}_L = \frac{\mathcal{E}}{\omega L}$ lagging on the applied

CHAPTER V.—PARA. 47

voltage by 90° , while the peak value of the current charging the condenser will be $\mathcal{I}_C = \omega C \mathcal{E}$, leading on the applied voltage by 90° . The peak value of the supply current $\mathcal{I}_s$ from the alternator will be the vector sum of the currents through the inductance and the condenser. In figs. 37, (b) and (c) are shown the different conditions which may obtain according to the frequency of the supply. If $\mathcal{I}_L$ is greater than $\mathcal{I}_C$ the supply current $\mathcal{I}_s$ will lag on the applied voltage by 90° , while if $\mathcal{I}_C$ is greater than $\mathcal{I}_L$ the supply current will lead on the applied voltage by 90° . It will be observed that as $\mathcal{I}_C$ and $\mathcal{I}_L$ are 180° out of phase with each other, the vector sum of these is actually their numerical difference; this may be written $\mathcal{I}_s = |\mathcal{I}_C - \mathcal{I}_L|$ the vertical lines indicating that the numerical value of the expression is denoted. As

$$\mathcal{I}_L = \frac{\mathcal{E}}{\omega L} \quad \mathcal{I}_C = \omega C \mathcal{E}, \quad \mathcal{I}_s = |\mathcal{I}_C - \mathcal{I}_L|,$$

$$\mathcal{I}_s = \mathcal{E} \left(\omega C - \frac{1}{\omega L} \right)$$

If the value of $\omega C - \frac{1}{\omega L}$ is positive, the current leads on the voltage; if negative the current lags.

Example 10.—An inductance of 160 microhenries and a capacitance of $\cdot 0002$ microfarad are connected in parallel, the losses in both components being negligible. If an E.M.F. of 100 volts R.M.S. at 800 kc/s is applied, find the R.M.S. supply current and whether lagging or leading.

$$I_s = E \left(\omega C - \frac{1}{\omega L} \right)$$

$$\begin{aligned} \omega C &= 6 \cdot 28 \times 800 \times 1,000 \times \cdot 0002 \times 10^{-6} \\ &= \cdot 001005 \end{aligned}$$

$$\begin{aligned} \frac{1}{\omega L} &= \frac{10^6}{6 \cdot 28 \times 800 \times 1,000 \times 160} \\ &= \frac{1}{804} \\ &= \cdot 001245 \end{aligned}$$

$$\begin{aligned} \omega C - \frac{1}{\omega L} &= \cdot 001005 - \cdot 001245 \\ &= - \cdot 00024 \end{aligned}$$

$$\begin{aligned} I_s &= 100 \times (- \cdot 00024) \\ &= - \cdot 024 \text{ amperes} \end{aligned}$$

Since I_s is negative, the current lags on the voltage by 90° . The effective reactance of the parallel combination is therefore inductive, and the value of the equivalent inductance is easily found.

$$X_e = \frac{E}{I_s}$$

$$L_e = \frac{E}{\omega I_s}$$

$$\begin{aligned} \frac{E}{I_s} &= \frac{100}{\cdot 024} \\ &= 4,160 \end{aligned}$$

$$\begin{aligned} L_e &= \frac{4,160}{6 \cdot 28 \times 800 \times 1,000} \times 10^6 (\mu H) \\ &= 830 \mu H \end{aligned}$$

Parallel resonance

48. In the particular case when $\omega C = \frac{1}{\omega L}$ the current $\mathcal{I}_0$ is equal to the current $\mathcal{I}_L$ and their arithmetical difference is zero, i.e. $\mathcal{I}_s = 0$. The circuit will therefore take no current whatever from the alternator, and the total effective impedance of the circuit is infinitely great. This condition is not possible in practice because both inductive and capacitive branches must possess some resistance, but may be closely approached. Assuming the resistance to be absolutely negligible, the frequency at which the supply current is zero is easily derived.

$$\begin{aligned}\text{When } \omega C &= \frac{1}{\omega L} \\ \omega^2 &= \frac{1}{LC} \\ \text{and } f &= \frac{1}{2\pi\sqrt{LC}}\end{aligned}$$

This is the resonant frequency of the circuit, which is said to be a rejector for this particular frequency. The conditions of parallel resonance are illustrated in fig. 38. Referring to the circuit diagram, at any instant when the current in the inductance is flowing downwards, the

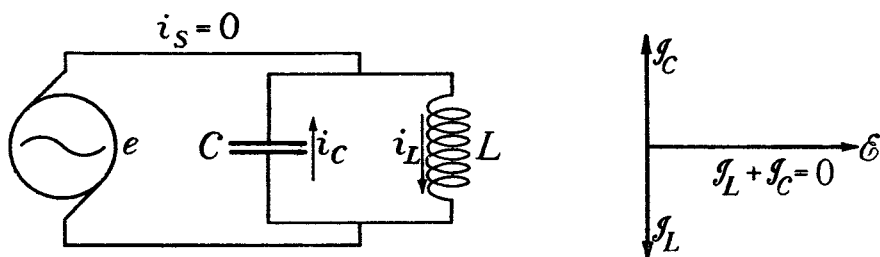


FIG. 38, CHAP. V.—Currents in circuit of Fig. 37, at resonant frequency.

current in the capacitance, which is 180° out of phase with it, must be flowing upwards, and the effect is that of an alternating current flowing to and fro in the closed circuit comprised by the inductance and condenser. For this reason, the current $I_L = I_c$ is often called the (R.M.S.) circulating current; its value depends upon the supply voltage and upon the ratio of capacitance to inductance.

$$\begin{aligned}\text{Since } I_L &= \frac{E}{\omega L} \text{ and } I_c = \omega C E \\ I_L \times I_c &= I_L^2 \text{ or } I_c^2 \\ I_L \times I_c &= \frac{E}{\omega L} \times \omega C E = \frac{C}{L} E^2 \\ \therefore I_L \text{ and } I_c &= E \sqrt{\frac{C}{L}}\end{aligned}$$

The current I_L is greater than I_c when $\frac{1}{\omega L}$ is greater than ωC , that is at frequencies below the resonant frequency. The supply current then lags by 90° on the applied voltage. On the other hand, at frequencies above resonance, ωC is greater than $\frac{1}{\omega L}$ and I_c is greater than I_L . The supply current will then lead on the applied voltage by 90° .

CHAPTER V.—PARA. 49

The effect of a small resistance in a rejector circuit

49. In practice, every circuit must contain some resistance however small, and it is now intended to consider the effect upon the action of a rejector circuit of a resistance R_L of the order which would be encountered in a practical, efficient circuit. Suppose the resistance to exist only in the inductive branch of the circuit. Remembering that $\chi_L = \frac{\omega L}{R_L} = \tan \theta_L$ where θ_L is the angle of lag of $\mathcal{I}_L$, reference to a table of tangents shows that if $\chi_L = 50$, $\mathcal{I}_L$ will lag on the applied voltage by very nearly 89° , while if $\chi_L = 100$, $\mathcal{I}_L$ will lag by nearly 89.5° . As the condenser is supposed to have no losses, the current $\mathcal{I}_C$ will lead on the applied voltage by 90° .

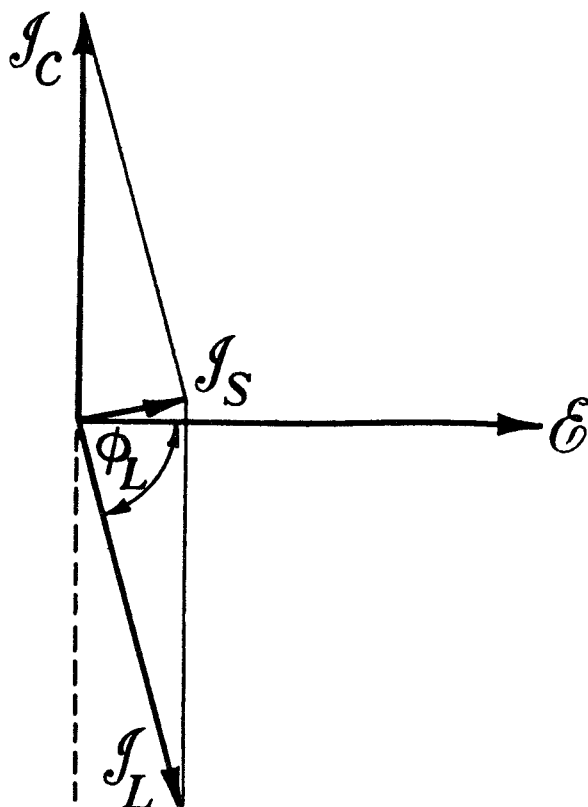


FIG. 39, CHAP. V.—Make-up current in rejector circuit possessing resistance.

$\mathcal{I}_L$ and $\mathcal{I}_C$ are thus practically 180° out of phase with each other, and at any frequency except the resonant frequency the R.M.S. value of the supply current ($\mathcal{I}_S$) is to all intents and purposes the difference between $\mathcal{I}_L \left(= \frac{\mathcal{I}_L}{\sqrt{2}} \right)$ and $\mathcal{I}_C \left(= \frac{\mathcal{I}_C}{\sqrt{2}} \right)$.

At the resonant frequency the power expended in the circuit will be $\mathcal{I}_S^2 R_L$ or $\frac{\mathcal{I}_S^2 R_L}{2}$ watts, but $\mathcal{I}_S$ is not the arithmetic difference between $\mathcal{I}_L$ and $\mathcal{I}_C$ because $\mathcal{I}_S$ is now practically in phase with the applied voltage. The vector diagram fig. 39 has been drawn to explain this, although the angle ϕ_L has been shown as much less than 89° for clearness. The value of the

supply current must depend upon the power expended in the circuit. Now this power is $I_L^2 R_L$ (or $I_C^2 R_L$ because $I_L^2 = I_C^2$) therefore the power expended is $I_L I_C R_L$ watts or

$$P = \frac{E}{\omega L} \times \omega C E \times R_L$$

$$= \frac{C R_L}{L} E^2 \text{ watts.}$$

As the supply current is in phase with the supply voltage, the circuit as a whole must be acting as a resistance. There is nothing new in this, for we saw that the acceptor circuit at its resonant frequency behaved as if it contained resistance only. Now in any circuit whatever the power expended may be expressed as $\frac{E^2}{R_d}$ where E is the R.M.S. value of the applied E.M.F. and R_d is the effective resistance of the circuit. The power $\frac{C R_L}{L} E^2$ may therefore be equated to $\frac{E^2}{R_d}$, giving the effective resistance of the circuit ;

$$\frac{E^2}{R_d} = \frac{E^2 C R_L}{L}$$

$$\therefore R_d = \frac{L}{C R_L}$$

The effective resistance of the rejector circuit at its resonant frequency is frequently referred to as its dynamic resistance. The R.M.S. value, I_s , of the supply current is $\frac{E}{R_d}$ or $\frac{E C R}{L}$ amperes, if E is in volts. It must be particularly borne in mind that the greater the actual resistance of the circuit the less is its dynamic resistance.

Resonance curves for a rejector circuit

50. Let us now take a rejector circuit containing a small resistance, and vary the frequency of the applied E.M.F. as was done in the study of an acceptor circuit, keeping its R.M.S. value constant, say 10 volts. Suppose that $L = 1.6 \mu H$, $C = .025 \mu F$, and $R = .064$ ohms. Then as the frequency is varied from 764 to 828 kc/s the corresponding variation of current is shown by curve (i) of fig. 40. Such a curve is called the resonance curve of the rejector circuit. It will be observed that at the resonant frequency, 796 kc/s, the current falls to a value $\frac{E C R}{L} = .01$ ampere, rising on each side of the resonant frequency. The rate at which the current increases, as the frequency is varied above or below the resonant frequency, depends upon the ratio $\frac{L}{R}$, as in the acceptor circuit. Thus if R remains constant, the effect of an alteration in the ratio $\frac{L}{C}$ is shown by curves (ii) and (iii). In curve (ii) R is .064 ohms but $L = 3.2 \mu H$ and $C = .0125 \mu F$, while in curve (iii) R is .064 ohms, $L = 1.13 \mu H$, $C = .0353 \mu F$. Now it will be observed that the current at any non-resonant frequency is greater in curve (iii) than in curve (i) or (ii) and it is therefore often erroneously concluded that the "selectivity" of a rejector circuit is increased as the ratio $\frac{C}{L}$ is increased. Before proceeding further, it must be pointed out that when speaking of a circuit consisting of an inductance and condenser in parallel, the term selectivity must be used with some caution. If the circuit is used as a true rejector, that is to suppress current at one particular frequency, then the criterion of its "selectivity" is the ratio (current passed by the device at the desired frequency) over (current passed at an undesired frequency). In fig 40 the undesired frequency would be 796 kc/s, while we may suppose that the desired frequency i.e. that which is required to pass through the rejector circuit, is 780 kc/s. Then in curve (i) the

CHAPTER V.—PARA. 51

ratio of desired (I_n) to undesired, (I_r), currents is 5.3 to 1, whereas in curve (ii) it is = 10 to 1, and in curve (iii) it is 3.6 to 1. The circuit corresponding to curve (ii) is thus the best of the three circuits as a "wave trap" as this form of current suppressor is generally called.

Now let us study the effect of varying R , while keeping L and C constant. Suppose that we take the original circuit, $L = 1.6 \mu H$, $C = .025 \mu F$ but reduce R to .032 ohms. Then the resulting resonance curve is shown in fig. 41 curve (ii), curve (i) being repeated from the previous figure to serve as a basis of comparison. It is again obvious that taking the ratio I_n/I_r , or (current at any non-resonant frequency) / (current at resonant frequency) as a criterion the selectivity of each circuit is proportional to the ratio $\frac{L}{R}$. The greater this ratio is, the greater the proportion of current at non-resonant frequency to the current at resonance.

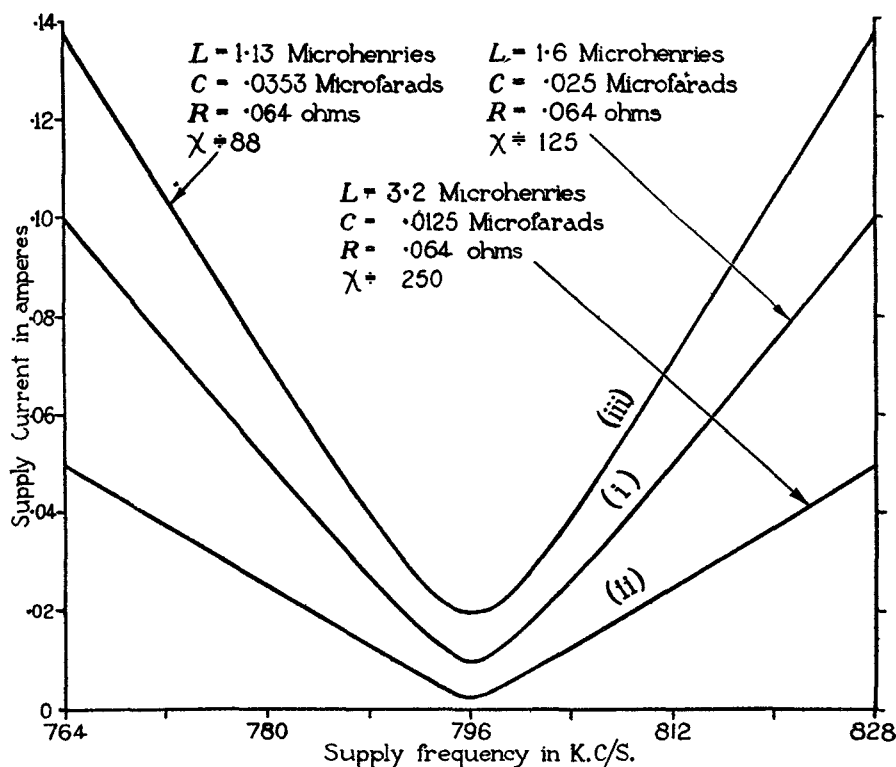


FIG. 40, CHAP. V.—Parallel resonance curves. Effect of ratio of inductance to capacitance.

When a rejector circuit is used in such a manner that the ratio of resonant to non-resonant P.D. at its terminals is of major importance, the above conclusions must be modified. The effect of any impedance which is effectively in parallel with it must be taken into consideration and it is desirable to analyse any particular case from first principles rather than to rely on general conclusions. An example of such an analysis is given in paragraph 64.

Circuit magnification of a rejector circuit

51. It should be obvious that a circuit consisting of L and C in parallel, whether inherent resistance is present or not, cannot have a voltage magnification, for the voltage across the inductance or condenser cannot be greater than the applied voltage. Instead the magnification

takes the form of current magnification that is, the circulating current is χ times as great as the supply current. This is easily shown as follows. The R.M.S. circulating current, is equal to $\frac{E}{\omega L}$ or $\omega C E$, while the supply current I_s is equal to $\frac{CR}{L} E$. The ratio I_L/I_s is therefore

$$\begin{aligned} & \frac{E}{\omega L} \div \frac{CR}{L} E \\ &= \frac{E}{\omega L} \times \frac{L}{CRE} = \frac{1}{\omega CR} = \chi \\ \text{or} \quad & \omega CE \times \frac{L}{CRE} = \frac{\omega L}{R} = \chi \end{aligned}$$

and the statement that $I_L = \chi I_s$ is proved. This relation can often be used to shorten work in connection with actual radio circuits.

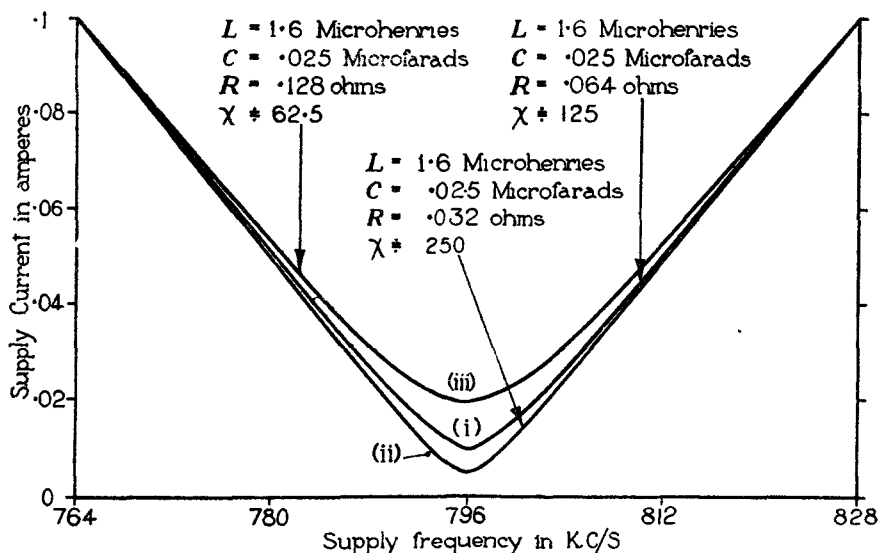


FIG. 41, CHAP. V.—Parallel resonance curves. Effect of variation of resistance.

Rejector circuit with considerable resistance

52. Although the circumstances rarely occur in radio circuits, it is desirable from a purely theoretical point of view to consider the conditions arising in a parallel inductance-capacitance combination when the resistance of either or both branches is comparable in magnitude with the reactance of the branch. Let us assume therefore that a circuit consists of an inductance of L henries which has a resistance of R_L ohms and a condenser of capacitance C farads, the losses in which can be represented by a resistance of R_C ohms. These being connected in parallel, an alternating E.M.F. of peak value $\mathcal{E}$ volts, and of variable frequency, is applied. The current in the inductance will be $\mathcal{I}_L = \frac{\mathcal{E}}{Z_L}$ where $Z_L = \sqrt{R_L^2 + (\omega L)^2}$ and will lag by an angle $\varphi_L = \tan^{-1} \frac{\omega L}{R_L}$

on the applied voltage. The current charging the condenser will be $\mathcal{I}_C = \frac{\mathcal{E}}{Z_C}$ where $Z_C = \sqrt{R_C^2 + \left(\frac{1}{\omega C}\right)^2}$ and will lead by an angle $\varphi_C = \tan^{-1} \frac{1}{\omega C R_C}$. The vector diagram of the

CHAPTER V.—PARA. 52

resulting conditions is given in fig. 42. The supply current $\mathcal{I}_s$ is given by the vector sum of $\mathcal{I}_L$ and $\mathcal{I}_C$, and is obtained by completing the parallelogram two sides of which are $\mathcal{I}_L$ and $\mathcal{I}_C$. The diagonal through the origin is then equal to the supply current.

In order to find an algebraic expression for the supply current, in terms of the applied E.M.F. and the circuit constants, it is necessary to resolve $\mathcal{I}_L$ and $\mathcal{I}_C$ each into two components in phase with, and 90° out of phase with the voltage. The components of $\mathcal{I}_L$ are (i) $\mathcal{I}_L \sin \phi_L$, 90° out of phase, and (ii) $\mathcal{I}_L \cos \phi_L$ in phase, with the applied E.M.F. If any doubt is felt as to the justification for this procedure, it may be dispelled by confirming that the vector sum of $\mathcal{I}_L \sin \phi_L$ and $\mathcal{I}_L \cos \phi_L$ is $\mathcal{I}_L$. Now the vector sum of these is $\sqrt{(\mathcal{I}_L \cos \phi_L)^2 + (\mathcal{I}_L \sin \phi_L)^2}$ and since $(\cos a)^2 + (\sin a)^2$ where a is an angle whatever, is equal to unity, we may consider the proposition proved. In the same way, the components of $\mathcal{I}_C$ are (i) $\mathcal{I}_C \sin \phi_C$, 90° out of phase, and (ii) $\mathcal{I}_C \cos \phi_C$ in phase, with the voltage.

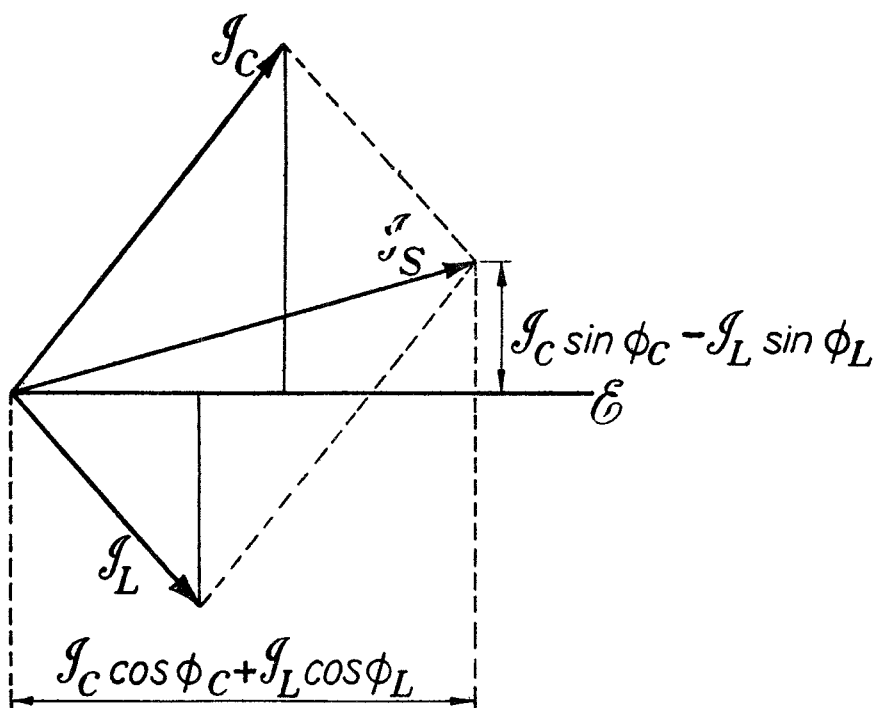


FIG. 42, CHAP. V.—Vector diagram showing relation between currents in rector circuit possessing considerable resistance.

Now consider the vector $\mathcal{I}_s$ representing the supply current. $\mathcal{I}_s$ can also be divided into in-phase and 90° out-of-phase components; its out-of-phase component is equal to the difference between $\mathcal{I}_L \sin \phi_L$ and $\mathcal{I}_C \sin \phi_C$, while its in-phase component is equal to the sum of $\mathcal{I}_L \cos \phi_L$ and $\mathcal{I}_C \cos \phi_C$. It should be noted that this is true even if the circuit possesses no resistance, for in this instance $\sin \phi_L$ and $\sin \phi_C$ are each equal to unity because both inductive and capacitive currents are 90° out of phase and $\sin 90^\circ = 1$, while $\cos \phi_L$ and $\cos \phi_C$ are each equal to zero because $\cos 90^\circ = 0$. Hence it may be said that in all cases

$$\mathcal{I}_s = \sqrt{(\mathcal{I}_L \sin \phi_L - \mathcal{I}_C \sin \phi_C)^2 + (\mathcal{I}_L \cos \phi_L + \mathcal{I}_C \cos \phi_C)^2}$$

In order to eliminate the trigonometrical terms from this equation we use the relation

$$\begin{aligned}\sin \varphi_L &= \frac{X_L}{Z_L} & \sin \varphi_C &= \frac{X_C}{Z_C} \\ \cos \varphi_L &= \frac{R_L}{Z_L} & \cos \varphi_C &= \frac{R_C}{Z_C}\end{aligned}$$

giving

$$\begin{aligned}Y_s &= \sqrt{\left(\frac{G}{Z_L} \times \frac{X_L}{Z_L} - \frac{G}{Z_C} \times \frac{X_C}{Z_C}\right)^2 + \left(\frac{G}{Z_L} \times \frac{R_L}{Z_L} + \frac{G}{Z_C} \times \frac{R_C}{Z_C}\right)^2} \\ &= G \sqrt{\left(\frac{X_L}{Z_L^2} - \frac{X_C}{Z_C^2}\right)^2 + \left(\frac{R_L}{Z_L^2} + \frac{R_C}{Z_C^2}\right)^2}\end{aligned}$$

The expression under the square root sign is the admittance of the parallel circuit. The susceptance of the combination is

$$\frac{X_L}{Z_L^2} - \frac{X_C}{Z_C^2} = B,$$

and the conductance is

$$\frac{R_L}{Z_L^2} + \frac{R_C}{Z_C^2} = G$$

The supply current will be in phase with the supply voltage when the susceptance of the circuit is zero. This occurs when

$$\frac{X_L}{Z_L^2} = \frac{X_C}{Z_C^2}$$

and the frequency at which this equation is true is called the resonant frequency of the circuit.

This frequency is found by expressing the equation in such a way that the frequency, or ω which is 2π times the frequency appears thus:—

$$\begin{aligned}\frac{\omega L}{R_L^2 + \omega^2 L^2} &= \frac{\frac{1}{\omega C}}{R_C^2 + \frac{1}{\omega^2 C^2}} \\ \frac{\omega L}{R_L^2 + \omega^2 L^2} &= \frac{\omega C}{\omega^2 C^2 R_C^2 + 1} \\ \frac{L}{R_L^2 + \omega^2 L^2} &= \frac{C}{\omega^2 C^2 R_C^2 + 1}\end{aligned}$$

cross multiplying,

$$\omega^2 L C^2 R_C + L = R_L^2 C + \omega^2 L^2 C$$

Collecting terms containing ω to the left hand side.

$$\omega^2 (L C^2 R_C^2 - L^2 C) = R_L^2 C - L$$

$$\text{whence } \omega^2 = \frac{R_L^2 C - L}{L C^2 R_C^2 - L^2 C}$$

$$= \frac{1}{L C} \left\{ \frac{R_L^2 C - L}{C R_C^2 - L} \right\}$$

$$\omega = \sqrt{\frac{1}{L C} \left\{ \frac{R_L^2 C - L}{C R_C^2 - L} \right\}}$$

$$\text{Finally } f = \frac{1}{2\pi} \sqrt{\frac{1}{L C} \left\{ \frac{R_L^2 C - L}{C R_C^2 - L} \right\}}.$$

CHAPTER V.—PARA. 53

From these equations it can be seen that only in the particular case when $R_c = R_L$ is the resonant frequency equal to

$$\frac{1}{2\pi\sqrt{LC}}$$

In most circumstances however, the factor enclosed in brackets only differs from unity by a very small fraction.

An interesting effect occurs if the numerator and denominator of the bracketed portion of the equation both become equal to zero. The "resonant frequency" according to the formula then becomes

$$f = \frac{1}{2\pi} \sqrt{\frac{1}{LC}} \times \frac{0}{0}$$

Now zero is not a number but the absence of any number whatsoever, and when an expression takes the form $\frac{0}{0}$ it is said to be indeterminate. The condition in which the resonant frequency becomes indeterminate is when $R_L = R_c = \sqrt{\frac{L}{C}}$. The circuit then behaves as if each branch has a resistance $\sqrt{\frac{L}{C}}$ in series with its inherent resistance, and the two branches in parallel then have a total joint resistance of $\sqrt{\frac{L}{C}}$ ohms, no matter what the applied frequency may be.

This phenomenon has an analogy in the case of a transmission line which contains inductance and capacitance distributed all along its length, its inductance per unit length being l henries and its capacitance per unit length being c farads. The line will then transmit all frequencies equally well if the "load" at the end of the transmission line consists of resistance only, its value being equal to $\sqrt{\frac{l}{c}}$ ohms.

53. Reverting to the consideration of the admittance Y , of a rejector circuit, we have seen that

$$Y = \sqrt{G^2 + B^2}$$

The latter expression can be considerably simplified for practical use with negligible sacrifice of accuracy, by introducing the figures of merit, χ_L , χ_c thus

$$\frac{X_L}{Z_L^2} = \frac{\omega L}{R_L^2 + \omega^2 L^2} = \frac{\omega L}{\omega^2 L^2 \left(1 + \frac{R_L^2}{\omega^2 L^2}\right)} = \frac{\chi_L^2}{\omega L (1 + \chi_L^2)}$$

$$\frac{X_c}{Z_c^2} = \frac{\frac{1}{\omega C}}{R_c^2 + \frac{1}{\omega^2 C^2}} = \frac{\omega C}{1 + \omega^2 C^2 R_c^2} = \frac{\omega C \chi_c^2}{1 + \chi_c^2}$$

$$\text{Hence } B = \omega C \left(\frac{\chi_c^2}{1 + \chi_c^2} \right) - \frac{1}{\omega L} \left(\frac{\chi_L^2}{1 + \chi_L^2} \right)$$

and when χ_L and $\chi_c \gg 1$, which is almost always the case in radio circuits, this simplifies to

$$B \doteq \omega C - \frac{1}{\omega L}.$$

Similarly

$$\frac{R_L}{Z_L^2} = \frac{1}{R_L \left(1 + \frac{\omega^2 L^2}{R_L^2}\right)} = \frac{1}{R_L (1 + x_L^2)}$$

$$\frac{R_C}{Z_C^2} = \frac{1}{R_C \left(1 + \frac{1}{\omega^2 C^2 R_C^2}\right)} = \frac{1}{R_C (1 + x_C^2)}$$

Hence

$$G = \frac{1}{R_L (1 + x_L^2)} + \frac{1}{R_C (1 + x_C^2)}$$

Again if x_L and $x_C \gg 1$ this simplifies to

$$G = \frac{1}{R_L x_L^2} + \frac{1}{R_C x_C^2},$$

and if R_C is negligible,

$$G = \frac{1}{R_L x_L^2} = \frac{R_L}{\omega^2 L^2}.$$

For nearly all practical purposes, then,

$$Y = \sqrt{\left(\omega C - \frac{1}{\omega L}\right)^2 + \left(\frac{1}{R_L x_L^2} + \frac{1}{R_C x_C^2}\right)^2}$$

At frequencies very near to the resonant frequency, the conductance may be considered equal to the conductance at resonance, that is $\frac{1}{R_d}$ or $\frac{CR}{L}$. The admittance is then

$$Y = \sqrt{\left(\omega C - \frac{1}{\omega L}\right)^2 + \left(\frac{CR}{L}\right)^2}$$

the R.M.S. supply current being found by the relation $I = YE$ as usual.

Example 11.—A coil having an inductance of 160 microhenries and a resistance of 19 ohms is connected in parallel with a condenser having a capacitance of .000256 microfarad and a resistance of 1 ohm. Obtain an approximation to the admittance at the frequency corresponding to $\omega = 5 \times 10^6$, the current set up by an R.M.S. voltage of 100 volts, and the angle of lag or lead.

$$\omega C = .000256 \times 10^{-6} \times 5 \times 10^6 = .00128$$

$$\frac{1}{\omega L} = \frac{1}{5 \times 10^6 \times 160 \times 10^{-6}} = \frac{1}{800} = .00125$$

$$B = \omega C - \frac{1}{\omega L} = .00003$$

Total resistance = 20 ohms.

$$G = \frac{CR}{L} = \frac{.000256 \times 20}{160} = .000032$$

$$Y = \sqrt{G^2 + B^2}$$

$$= \sqrt{(32 \times 10^{-6})^2 + (30 \times 10^{-6})^2}$$

$$= 10^{-6} \sqrt{1,024 + 900}$$

$$= 43.8 \times 10^{-6} \text{ siemens (or mho)}$$

$$I = YE = 43.8 \times 10^{-6} \times 100$$

$$= 4,380 \text{ microamperes.}$$

CHAPTER V.—PARA. 54

As ωC is greater than $\frac{1}{\omega L}$ the current leads on the applied voltage by an angle $\theta = \tan^{-1} \frac{B}{G}$

$$\frac{B}{G} = \frac{.00003}{.000032} = .936$$

$$\therefore \theta \doteq 43^\circ$$

VECTOR OPERATORS

54. A stage has now been reached at which the relations between current voltage and impedance in various types of A.C. circuit can be found by the use of vector diagrams in conjunction with elementary trigonometry. It is possible to solve practically any A.C. problem without further mathematical knowledge, but many problems are much more easily handled by expressing the vector quantities themselves by symbols denoting not only their magnitude but also their direction. One method which is sometimes employed in connection with impedances is to denote the magnitude of the impedance by a symbol or by a number representing the magnitude followed by a symbol or number representing the angle of phase difference caused by this impedance, for example, (Z, θ) would be an impedance of magnitude Z , causing a phase difference of θ radians, but as this alone would not indicate whether the impedance caused a leading or lagging current, the sign $\angle$ is used to denote an angle of lag, and ∇ to denote an angle of lead. Thus $Z \angle \theta$ is equivalent to $\sqrt{R^2 + \omega^2 L^2}$, where $\frac{\omega L}{R} = \tan^{-1} \theta$.

$600 \angle \frac{\pi}{2} = \sqrt{R^2 + \omega^2 L^2}$, where $\frac{\omega L}{R} = \tan^{-1} \frac{\pi}{2}$; but $\tan^{-1} \frac{\pi}{2} = \infty$, and $\frac{\omega L}{R}$ is consequently infinite also. As ωL cannot be infinite, because $\sqrt{R^2 + \omega^2 L^2}$ is only 600 ohms, R must be zero, and therefore $600 \angle \frac{\pi}{2}$ represents an inductive reactance of 600 ohms. A further example may be given: let $Z \angle \theta = 800 \angle .5326$, or $800 \angle 30^\circ$ $.5326$ being the radian measure of 30° .

$$\text{Then } \sqrt{R^2 + \omega^2 L^2} = 800$$

$$\text{and } \frac{\omega L}{R} = \tan 30^\circ = .5774 \text{ or } \frac{1}{\sqrt{3}}$$

$$\frac{\omega^2 L^2}{R^2} = \frac{1}{3}$$

$$\omega^2 L^2 = \frac{R^2}{3}$$

$$\text{Also } R^2 + \omega^2 L^2 = 6,400$$

$$R^2 \left(1 + \frac{1}{3}\right) = 6,400$$

$$R^2 = \frac{6,400 \times 3}{4} = 4,800$$

$$\therefore R = 69.2 \text{ ohms,}$$

$$\text{while } \omega^2 L^2 = \frac{4,800}{3} = 1,600$$

$$\text{and } \omega L = 40 \text{ ohms.}$$

This method of indicating the nature of an impedance of a given magnitude is frequently employed in telephone engineering.

55. A method which leads itself more readily to manipulation is that which introduces the conception of a "vector operator." In order to explain this, it may be desirable to consider the idea of a mathematical operation. When a quantity is denoted by $a + b$, the plus sign may be considered to denote an operation performed upon b , for we may consider that originally there were two groups containing a objects and b objects respectively, and an operation is performed upon the latter group, the whole being conveyed to and possibly intermingled with the former group so that a new group is formed, containing a number of objects equal to the sum of the numbers in the two original groups. If instead of $a + b$, we write $b + a$, we may consider that the group denoted by b is stationary and that the operation of addition is performed by moving the group a . The numerical result is of course the same in each case, and $a + b = b + a$.

The plus sign thus indicates the operation of addition, while the minus sign when used in arithmetic signifies the operation of subtraction, e.g. $a - b$ means that from a group of a objects a number b is taken. The minus sign is also called the negative sign, for it is also used to convey the conception of contrariety of direction with regard to a given direction which is arbitrarily said to be positive. The two meanings of the signs $+$ and $-$ rarely if ever give rise to confusion. If a certain problem concerns objects, the equation $7 - 3 = 4$, for example, indicates that the operation of subtraction is performed upon three members of a group of seven, and the residue will consist of four members. The expression $3 - 7 = -4$ is meaningless when applied to objects, but is intelligible when applied to distances in space if movement in a certain direction is assumed to be positive and movement in a direction exactly opposite to this to be negative, for the equation may then be considered to signify that, starting from a point which may be denoted by zero (0) an object is moved through three units of distance in the positive direction, and afterwards through seven units in the contrary or negative direction, its final position being four units from the original position (or origin) in the negative direction.

56. Two symbols used in conjunction with numbers may also require definition. The symbol 0 (zero) signifies the absence of any number, while the symbol ∞ (infinity) is used to denote a whole series of numbers which are greater than it is possible to comprehend. It must not be thought that the sign ∞ denotes a single, extremely large number. The untruth of this can be seen by allowing ∞ to represent some number too great to admit of comprehension, and then considering it to be raised to the power n , where n is a number greater than unity. The result of this operation is a number larger than the original one, which was denoted by ∞ , but as it is too large for comprehension we still denote it by ∞ , which thus becomes a symbol denoting a range of numbers and not a single number. This leads to another conception of zero, for if infinity is defined as some number larger than can be conceived $\frac{1}{\infty}$ must be some number smaller

than admits of conception, and this exceedingly small quantity $\frac{1}{\infty}$ is denoted by zero, hence the sign 0 has a dual existence, sometimes signifying the entire absence of any number and sometimes a quantity smaller than it is possible to imagine. We see, therefore, that many of the signs used in ordinary arithmetic and algebra are capable of different interpretations, yet it is rare that any confusion arises as to their meaning in any given example.

When applied to distances in space, the symbol $+\infty$ represents an infinitely large distance in the positive direction, and $-\infty$ some infinitely large distance in the opposite direction. The whole series of numbers from $+\infty$ through zero to $-\infty$ with the exception of zero itself, are termed real numbers. This qualification "real" arose from the impossibility of representing the square root of a negative quantity by a number either positive or negative, because no quantity when multiplied by itself will give a negative quantity. Quantities like $\sqrt{-4}$, $\sqrt{-b}$, etc., were therefore called imaginary to distinguish them from quantities like $\sqrt{4}$, $\sqrt{6.28}$, etc., the square root of which can be determined with an error smaller than any assignable magnitude if sufficient decimal places are calculated. Any imaginary quantity can be expressed as the product of $\sqrt{-1}$ and a real quantity e.g.

$$\sqrt{-4} = \sqrt{(-1) \times 4} = \sqrt{-1} \times \sqrt{4} = \sqrt{-1} \times 2$$

CHAPTER V.—PARA. 57

The term $\sqrt{-1}$ is generally denoted in mathematical textbooks by the symbol i , but in electrical literature by the symbol j to avoid confusion with the symbol for current. The choice of the term "imaginary" to denote a quantity such as $\sqrt{-1} \times 2$ or $j2$ is unfortunate because it may lead to the impression that no meaning can be assigned to such a quantity, whereas it will be shown that " j " may be regarded as a symbol of operation.

57. We have seen that real numbers can be represented by distances from the origin along an arbitrary axis. Conventionally, the positive direction of this axis extends for an infinite distance to the right of the origin, while the negative direction of this axis extends for an infinite distance to the left of the origin. The multiplication of a positive quantity by -1 can therefore be considered as its reversal in direction, by rotation through 180° or π radians in either a clockwise or anticlockwise direction. Again, the result of multiplying $+a$ by $-b$ is the quantity $-ab$, which can be regarded as a multiplication of $+a$ by $+b$, giving a quantity of magnitude ab , and a rotation through 180° as in the previous instance. A further multiplication by -1 may be regarded as a second rotation of 180° , so that -1×-1 is equivalent to a single rotation of 360° or 2π radians. An operator which rotates a vector quantity in this way is termed a versor.

The quantity $\sqrt{-1}$ is a number which if multiplied by itself gives -1 . The application of $\sqrt{-1}$ twice in succession to the quantity a gives $\sqrt{-1} \sqrt{-1} a$ or $-a$, and is equivalent to a rotation of a through 180° or two right-angles. Now let us postulate an operator which will effect a rotation through only one right-angle; this operator may be termed a quadrantal versor. For convenience we may also define a unit vector as one having unit length and lying in the positive direction along the axis of real numbers. Then operating with the quadrantal versor upon the unit vector twice in succession will result in turning the latter through two right angles, which is π radians or 180° , that is two successive operations by the quadrantal versor are equivalent to a single multiplication by -1 , or by two successive multiplications by $\sqrt{-1}$, because $\sqrt{-1} \times \sqrt{-1} = -1$. The result of a single operation by the quadrantal versor therefore appears to be identical with that achieved by a multiplication by the factor $\sqrt{-1}$, or j . By convention the positive direction of rotation is anticlockwise. If the operation denoted by j is described as "jaying" the vector, we see that commencing with a unit vector, jaying the latter once results in a rotation through 90° , jaying twice a rotation through 180° , because $j \times j = j^2 = -1$. A further operation gives $j \times j \times j = j^3 = -1 \times j$ or $-j$, while a fourth gives $j \times j \times j \times j = j^4 = -1 \times -1 = +1$, and the unit vector becomes a unit vector once more,

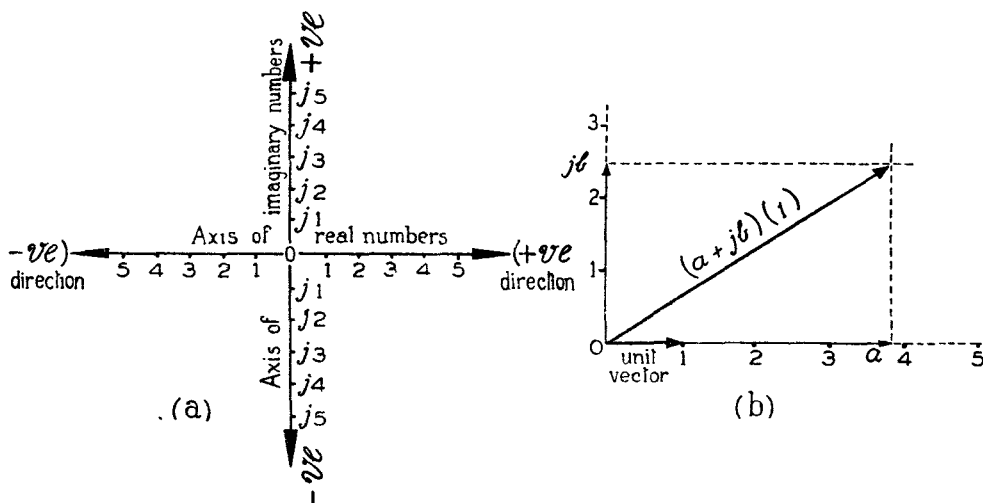


FIG. 43, CHAP. V.—Argand's diagram; algebraic representation of vectors.

having its original length and direction. With reference to the axis of real quantities the axis of a quantity upon which j has operated once only, is vertically upward from the origin, while the axis of those upon which $-j$ has operated once only is vertically downward through the origin. This may be shown upon what is known as Argand's diagram (fig. 43a). The fact that $\sqrt{-1}$ appears to lead a dual existence, sometimes appearing merely as an imaginary number and at others as a quadrantal versor need cause no confusion.

58. Complex numbers are those which consist of the sum of a real and imaginary number, e.g. $a + jb$. If two such quantities are equal to each other, e.g.

$$a + jb = c + jd$$

the real parts a and c must be equal to each other and the imaginary parts jb and jd also equal, because it is impossible for an imaginary number to be equal to a real number by definition.

Thus if $a + jb = c + jd$

$$\begin{aligned}(a + jb) - (a + jb) &= c + jd - (a + jb) \\ 0 &= c - a + j(d - b) \\ \therefore c &= a \text{ and } d = b\end{aligned}$$

unless $a - c = j(d - b)$, i.e. a real number equal to an imaginary one, which is impossible.

A complex number can be an operator. Suppose that the number $(a + jb)$ operates upon a unit vector, which may be denoted by (1). Then $(a + jb)(1) = a(1) + jb(1)$. The result of the operation is the sum of two vectors one of which is in the direction of unit vector but is a times the magnitude of the latter, while the other is b times the magnitude of the unit vector but has been "jayed" i.e. rotated through 90° in the anticlockwise direction. The magnitude of the vector $(a + jb)(1)$ is seen from its diagrammatic representation (fig. 43b) to be $\sqrt{a^2 + b^2}$. The angle through which the vector has rotated with reference to the direction of the unit vector is $\tan^{-1} \frac{b}{a}$. When used in conjunction with vector operators it is usual to denote vector quantities

by clarendon type. The original vector operated upon may be an alternating current, e.g. $\mathbf{I}$, and the operator $R + j\omega L$. Then $(R + j\omega L)\mathbf{I}$ is a vector which may be regarded as the sum of two vectors, viz. $R\mathbf{I} + j\omega L\mathbf{I}$ the first being an E.M.F. which is in phase with the current and the second an E.M.F. which leads on the current by 90° . The magnitude of the sum of these is $\sqrt{R^2 + \omega^2 L^2} I$ where I is the magnitude of the current $\mathbf{I}$, and the above result may be written symbolically as

$$e = Z \mathbf{I}$$

$$\text{or } e = (R + j\omega L) \mathbf{I}$$

When used in this way the complex number $R + j\omega L$ is called an impedance operator in order to distinguish it from the magnitude of the impedance viz. $\sqrt{R^2 + \omega^2 L^2}$, the latter often being called simply the impedance. Some writers define $R + j\omega L$ as the vector impedance and $\sqrt{R^2 + \omega^2 L^2}$ as the scalar impedance. It must be remembered however that $R + j\omega L$ is not itself a vector but a vector operator. The preceding results are often conveniently thrown into another form in which the point hitherto denoted by 0 is regarded as the origin of a system of polar co-ordinates, the length and direction of the vector whose magnitude is r being denoted by r, θ . If a vector is represented by $(a + jb)(1)$ where (1) is a unit vector, its magnitude or size is $\sqrt{a^2 + b^2} = r$, where $a = r \cos \theta$, $b = r \sin \theta$, and therefore

$$(a + jb)(1) = r(\cos \theta + j \sin \theta)(1)$$

hence the expression $r(\cos \theta + j \sin \theta)$ is equivalent to the operator $a + jb$. The factor r in this expression is equal to $\sqrt{a^2 + b^2}$ and is called the modulus of $a + jb$, while the angle θ which is equal to $\tan^{-1} \frac{b}{a}$ is called the argument of the complex quantity.

CHAPTER V.—PARAS. 59-60

59. Complex numbers can be multiplied together in the same way as ordinary numbers, thus $(a + jb)(c + jd) = ac + jad + jbc + j^2bd$, or (since $j^2 = -1$) $(a + jb)(c + jd) = (ac - bd) + j(ad + bc)$.

To divide one complex number by another an artifice is employed, e.g. $\frac{a + jb}{c + jd}$ is found by first expressing the denominator as a real quantity. To do this another complex number is introduced into the equation, namely $c - jd$, so that the quantity becomes

$$\frac{a + jb}{c + jd} \times \frac{c - jd}{c - jd}$$

$c - jd$ and $c + jd$ are called conjugate complex numbers. $(c + jd)(c - jd) = c^2 - jcd + jcd - j^2d^2$ or $c^2 + d^2$, because $-j^2 = 1$.

$$\begin{aligned} \text{Hence } \frac{a + jb}{c + jd} &= \frac{ac - jad + jbc - j^2bd}{c^2 + d^2} \\ &= \frac{ac + bd}{c^2 + d^2} + j \frac{bc - ad}{c^2 + d^2} \\ &= A + jB \end{aligned}$$

which is a new complex number. The process of expressing the quotient of two complex quantities as a single complex quantity in this way is known as rationalisation.

Demoivre's theorem

60. We have seen that if the operator $r(\cos \theta + j \sin \theta)$ operates upon a unit vector it has an effect equal to $a + jb$ if $a = r \cos \theta$ and $b = r \sin \theta$. Similar results are obtained if the operation is performed upon any vector, for example if

$$\mathbf{u} = (\cos \theta + j \sin \theta) \mathbf{v}$$

where $\mathbf{v}$ is a vector, and $\mathbf{u}$ is the result of operation upon it, the vector $\mathbf{u}$ is equal in magnitude to the vector $\mathbf{v}$ but is rotated in an anti-clockwise direction through the angle $\tan^{-1} \frac{b}{a}$. The inverse operation denoted by

$$\mathbf{v} = \frac{\mathbf{u}}{\cos \theta + j \sin \theta}$$

means that $\mathbf{v}$ is equal in magnitude to $\mathbf{u}$ but is rotated through the angle θ in a clockwise direction. Now, by the process of rationalisation

$$\begin{aligned} \frac{1}{\cos \theta + j \sin \theta} &= \frac{\cos \theta - j \sin \theta}{\cos^2 \theta + \sin^2 \theta} \\ &= \cos \theta - j \sin \theta, \text{ because } \cos^2 \theta + \sin^2 \theta = 1 \end{aligned}$$

As $\cos \theta = \cos(-\theta)$ and $-\sin \theta = \sin(-\theta)$, $\cos \theta + j \sin(-\theta)$ or $\cos \theta - j \sin \theta$ is an operator which causes clockwise rotation. Successive rotations, say of θ and ϕ , are equivalent to a single rotation $\theta + \phi$. That is to say the operation denoted by the expression

$\cos(\theta + \phi) + j \sin(\theta + \phi)$ is the effect of two successive rotations, the first being $(\cos \theta + j \sin \theta)$ and the second $(\cos \phi + j \sin \phi)$. Hence

$$\cos(\theta + \phi) + j \sin(\theta + \phi) = (\cos \theta + j \sin \theta)(\cos \phi + j \sin \phi)$$

If $\theta = \phi$

$$\cos 2\theta + j \sin 2\theta = (\cos \theta + j \sin \theta)^2$$

and if instead of only $\theta + \theta$, we have a sum of n rotations each equal to θ :—

$$\theta + \theta + \theta \dots \text{etc. } n \text{ times} = n\theta$$

$\cos n\theta + j \sin n\theta = (\cos \theta + j \sin \theta)^n$. From this important theorem many useful results can be derived, for example the addition formulae of trigonometry which are frequently used in alternating current theory, particularly in the consideration of modulated waves :—

$$\begin{aligned}\cos(a+b) + j \sin(a+b) &= (\cos a + j \sin a)(\cos b + j \sin b) \\ &= \cos a \cos b + j \cos a \sin b + j \sin a \cos b - \sin a \sin b\end{aligned}$$

Equating the real and imaginary parts

$$\begin{aligned}\cos(a+b) &= \cos a \cos b - \sin a \sin b \\ \sin(a+b) &= \cos a \sin b + \sin a \cos b\end{aligned}$$

If instead of b , we write $-b$, in the original equation, we obtain

$$\begin{aligned}\cos(a-b) &= \cos a \cos b + \sin a \sin b \\ \sin(a-b) &= \sin a \cos b - \cos a \sin b\end{aligned}$$

While if $a = b$,

$$\begin{aligned}\cos 2a + j \sin 2a &= (\cos a + j \sin a)^2 \\ &= (\cos^2 a + 2j \sin a \cos a - \sin^2 a) \\ \text{Hence } \cos 2a &= \cos^2 a - \sin^2 a \\ \text{and } \sin 2a &= 2 \sin a \cos a\end{aligned}$$

adding $\sin^2 a - \sin^2 a$ (which is zero) to the right-hand side of the first of the two previous results

$$\begin{aligned}\cos 2a &= \cos^2 a + \sin^2 a - 2 \sin^2 a \\ &= 1 - 2 \sin^2 a\end{aligned}$$

$$\text{because } \cos^2 a + \sin^2 a = 1$$

Rearranged this becomes

$$\sin^2 a = \frac{1}{2} (1 - \cos 2a).$$

a result we have already used in dealing with the average value of a sinusoidal curve over a complete period. Reverting to a former pair of expressions,

$$\cos(a+b) - \cos(a-b) = \cos a \cos b - \sin a \sin b - \cos a \cos b + \sin a \sin b$$

or

$$\cos(a-b) - \cos(a+b) = 2 \sin a \sin b$$

As a practical example of the use of these formulae, consider the expression

$$i = \mathcal{J} (1 + \sin \omega_a t) \sin \omega_r t,$$

where ω_a is 2π times an audio-frequency f_a and ω_r is 2π times a radio-frequency f_r . In Chapter XII it is shown that the aerial current in an R/T transmitter may be of this form.

Expanding the right-hand member

$$\begin{aligned}i &= \mathcal{J} [\sin \omega_r t + \sin \omega_r t \sin \omega_a t] \\ &= \mathcal{J} [\sin \omega_r t + \frac{1}{2} \{ \cos (\omega_r - \omega_a) t - \cos (\omega_r + \omega_a) t \}] \end{aligned}$$

so that the complicated expression first given may be resolved into the sum of three sinusoidal waves of different frequencies.

61. In order to develop still another method of representation of a vector quantity by algebraic symbols, we must first refer to what are termed algebraic series. The latter expression is used to denote a number of terms each of which is related to its predecessor and successor in a perfectly definite manner. For instance, 1, 2, 3, 4, 5, is a series, each term being formed by adding unity to the preceding one, while 1, 2, 4, 8, etc. is a series in which each term is formed by multiplying the preceding one by two. An example of the development of a series is found in

CHAPTER V.—PARA. 61

the algebraic expression $(a + x)^n$, which gives rise to what is known as the binomial series. The binomial theorem states that, no matter what values a , x and n possess, i.e., positive or negative, integral or fractional,

$$(a + x)^n = a^n + \frac{n}{1} \cdot a^{n-1} \cdot x + \frac{n}{1} \cdot \frac{n-1}{2} \cdot a^{n-2} \cdot x^2 \\ + \frac{n}{1} \cdot \frac{n-1}{2} \cdot \frac{n-2}{3} \cdot a^{n-3} \cdot x^3 \dots\dots\dots$$

This is easily proved when n is a small positive integer, e.g. 2, 3, or 4, by direct multiplication;

$$(a + x)^2 = a^2 + \frac{2}{1} \cdot a^{2-1}x + \frac{2}{1} \cdot \frac{2-1}{2} \cdot a^{2-2} \cdot x^2 \\ + \frac{2}{1} \cdot \frac{2-1}{2} \cdot \frac{2-2}{3} \cdot a^{2-3} \cdot x^3 \dots\dots\dots \\ = a^2 + 2a^1x + \frac{2}{1} \cdot \frac{1}{2} \cdot a^0 \cdot x^2 \\ + 2 \cdot \frac{1}{2} \cdot \frac{0}{3} \cdot a^{-1} \cdot x^3 \dots\dots\dots \\ = a^2 + 2ax + x^2$$

because $a^0 = 1$ and the fourth term of the series and all subsequent terms contain 0 as a factor and are therefore equal to zero.

Similarly

$$(a + x)^3 = a^3 + 3a^2x + 3ax^2 + x^3$$

The arrangement of the terms can be seen to follow a definite plan. Starting with a^n , the powers of a decrease by unity in each successive term until (if n is a positive integer) a^0 or unity is reached, when the series ceases. This, however, does not occur if n is fractional or negative, for no term ever contains a^0 , consequently in such instances the series continues indefinitely. The powers of x increase as the powers of a diminish, the first term being $a^n x^0$ or a^n , the second containing x^1 , the third x^2 and so on. The numerical coefficients can easily be written, provided n is a positive integer, by the use of the following table.

																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																					</
--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	----

It will be observed that any integer in this triangle is the sum of the two adjacent to it in the line above, except at the ends of each line where the integer is always unity. By adding the next line in the table for instance we may readily obtain the expansion of $(a + x)^8$

$$(a + x)^8 = a^8 + 8a^7x + 28a^6x^2 + 56a^5x^3 + 70a^4x^4 + 56a^3x^5 + 28a^2x^6 + 7ax^7 + x^8$$

62. Let the expression $\left(1 + \frac{1}{n}\right)^n$ be expanded by the binomial theorem. The series obtained is

$$\begin{aligned} 1^n + \frac{n}{1} \cdot 1^{n-1} \cdot \frac{1}{n} + \frac{n}{1} \cdot \frac{n-1}{2} \cdot 1^{n-2} \cdot \left(\frac{1}{n}\right)^2 \\ + \frac{n}{1} \cdot \frac{n-1}{2} \cdot \frac{n-2}{3} \cdot 1^{n-3} \cdot \left(\frac{1}{n}\right)^3 \dots \dots \text{etc.} \\ = 1 + 1 + \frac{n(n-1)}{1 \times 2 \times n^2} + \frac{n(n-1)(n-2)}{1 \times 2 \times 3 \times n^3} + \frac{n(n-1)(n-2)(n-3)}{1 \times 2 \times 3 \times 4 \times n^4} \dots \dots \\ = 1 + 1 + \frac{1}{2} + \frac{\left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right)}{2 \times 3} + \frac{\left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right)\left(1 - \frac{3}{n}\right)}{2 \times 3 \times 4} \dots \dots \end{aligned}$$

Now suppose n to become larger and larger until it is some quantity greater than is comprehensible, which we may denote by the sign ∞ . All the fractions of the form $\frac{2}{n}, \frac{3}{n}$ etc., become utterly insignificant when added to or subtracted from unity, and it may then be stated that the limiting value of $\left(1 + \frac{1}{n}\right)^n$, when $n \rightarrow \infty$ (or when $\frac{1}{n} \rightarrow 0$) is

$$1 + 1 + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!} + \frac{1}{5!} + \frac{1}{6!} \dots \dots \dots \text{etc.}$$

where $2! = 1 \times 2, 3! = 1 \times 2 \times 3$ etc., $3!$ is called "factorial 3", $n!$ "factorial n " and so on. The sign $\rightarrow$ is read "approaches the limiting value".

The sum of an infinitely large number of terms of this series is $2.718281828 \dots$ which is denoted by the greek letter e . The value of " e " to five correct decimal places can be obtained by taking only ten terms of the series. This awkward-looking number is of great importance in physics, being connected with all natural processes of growth or decay, for instance the voltage to which a condenser is charged by an applied steady voltage E is

$$e_t = E \left(1 - \frac{1}{e^{\frac{t}{CR}}} \right)$$

which is more conveniently written

$$e_t = E \left(1 - \frac{1}{e^{\frac{t}{CR}}} \right)$$

where e_t is the voltage t seconds after the charge commences, C the capacitance of the condenser, R the total resistance in series with it, and E the applied E.M.F.

Now suppose that the expression to be expanded, (that is, to be expressed as the sum of a series of terms) is $\left(1 + \frac{1}{n}\right)^{nx}$. This becomes, by employment of the binomial theorem

$$\begin{aligned} \left(1 + \frac{1}{n}\right)^{nx} = 1 + \frac{nx}{n} + \frac{nx(n-1)}{2! n^2} \\ + \frac{nx(n-1)(n-2)}{3! n^3} \dots \dots \dots \text{etc.} \end{aligned}$$

CHAPTER V.—PARA. 63

and by sufficiently increasing n , as in the previous discussion, the terms $nx - 1$, $nx - 2$ etc., can be made to differ inappreciably from nx . When $n \rightarrow \infty$, therefore this series becomes

$$\left\{ \left(1 + \frac{1}{n} \right)^n \right\}^x = e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} \dots \dots \dots \text{etc.}$$

In the above example

$$e^{\frac{t}{CR}} = 1 + \frac{t}{CR} + \frac{\left(\frac{t}{CR} \right)^2}{2!} + \frac{\left(\frac{t}{CR} \right)^3}{3!} \dots \dots \dots \text{etc.}$$

If x is a negative quantity, it can be shown in the same way that

$$e^{-x} = 1 - x + \frac{x^2}{2!} - \frac{x^3}{3!} + \frac{x^4}{4!} \dots \dots \dots \text{etc.}$$

Thus

$$e^{-\frac{t}{CR}} = 1 - \frac{t}{CR} + \frac{\left(\frac{t}{CR} \right)^2}{2!} - \frac{\left(\frac{t}{CR} \right)^3}{3!} \dots \dots \dots \text{etc.}$$

Exponential form of $\cos \theta + j \sin \theta$.

63 As n rotations through an angle of θ radians are equal to a single rotation of $n\theta$ radians

$$\begin{aligned} \cos n\theta + j \sin n\theta &= (\cos \theta + j \sin \theta)^n \\ &= \cos^n \theta \left(1 + j \frac{\sin \theta}{\cos \theta} \right)^n \\ &= \cos^n \theta (1 + j \tan \theta)^n \end{aligned}$$

If $n\theta = \varphi$

$$\cos \varphi + j \sin \varphi = \cos^n \frac{\varphi}{n} \left(1 + j \tan \frac{\varphi}{n} \right)^n,$$

and if n is allowed to become larger and larger without limit $\tan \frac{\varphi}{n}$ and $\frac{\varphi}{n}$ become more nearly equal to each other. In the limit, $\tan \frac{\varphi}{n} = \frac{\varphi}{n}$ and

$$\cos \varphi + j \sin \varphi = \cos^n \frac{\varphi}{n} \left(1 + j \frac{\varphi}{n} \right)^n$$

or, since the term $\cos^n \frac{\varphi}{n}$ also approaches unity as n is increased without limit

$$\cos \varphi + j \sin \varphi = \left(1 + j \frac{\varphi}{n} \right)^n \text{ if } n \rightarrow \infty.$$

The right hand member of this equation may be expanded by the binomial theorem, with the following result ;

$$\cos \varphi + j \sin \varphi = 1 + j\varphi + \frac{(j\varphi)^2}{2!} + \frac{(j\varphi)^3}{3!} + \frac{(j\varphi)^4}{4!} \dots \dots \text{etc.}$$

Comparing the righthand member with the expansion of e^x it will be seen that they are of the same form, and therefore the operator $(\cos \varphi + j \sin \varphi)$ is written in the alternative form $e^{j\varphi}$. It can also be shown that $\cos \varphi - j \sin \varphi$ is formally equivalent to $e^{-j\varphi}$.

It has already been shown that

$$(\cos \theta + j \sin \theta) (\cos \varphi + j \sin \varphi) = \cos (\theta + \varphi) + j \sin (\theta + \varphi)$$

and in the exponential form this becomes $e^{j\theta} e^{j\varphi} = e^{j(\theta + \varphi)}$. The imaginary index therefore enters into algebraic combination just as if it were a real number.

The four methods of expressing a vector operator are equivalent to each other, and if $Z \angle \theta$, $a + jb$, $Z (\cos \theta + j \sin \theta)$, $e^{j\theta}$ are applied in turn to a unit vector, the effect of the operation is in every case to rotate the vector in the positive direction through the angle θ and to extend the magnitude of the vector to Z units, provided that $a^2 + b^2 = Z^2$ and $\theta = \tan^{-1} \frac{b}{a}$.

Example 12.— $3 + j4$ is the $a + jb$ form of a vector operator. Express in the other forms.

The modulus of $3 + j4$ is $\sqrt{3^2 + 4^2} = \sqrt{25} = 5$

The argument, θ , is $\tan^{-1} \frac{4}{3} = .927$ radians approx.

$$3 + j4 = 5 (\cos .927 + j \sin .927)$$

$$\cos .927 = .6$$

$$\sin .927 = .8$$

$$3 + j4 = 5 (.6 + j .8)$$

$$\text{or } 5 \angle .927$$

$$\text{or } 5 \angle 53^\circ 7'$$

$$\text{or } 5 e^{j.927}$$

The advantage of the form $e^{j\phi}$ is the manner in which it lends itself to multiplication and division of operators. This process is simply carried out as follows:—

$$\text{if } a + jb = Z_1 e^{j\phi}$$

$$c + jd = Z_2 e^{j\theta}$$

$$(a + jb)(c + jd) = Z_1 Z_2 e^{j(\theta + \phi)}$$

While

$$Z_1 e^{j\phi} \div Z_2 e^{j\theta} \text{ is } \frac{Z_1}{Z_2} e^{j(\phi - \theta)}$$

The exponential form $Z e^{j\phi}$ is therefore the most convenient when multiplication or its extensions are to be performed, while the form $a + jb$ is preferable when addition or subtraction of vector operators is contemplated. For this reason it is often advisable to change from one form to the other in the course of an analysis.

64. As an example of the simplification introduced into A.C. calculations by the employment of the impedance operator, the selectivity of the circuit consisting of an inductance and capacitance in parallel will now receive further consideration. In the first place, it must be emphasised that this parallel combination (which for the sake of brevity may be referred to as a "rejector circuit", even though the particular application is not that of a true rejector), is only of practical utility when used in combination with other circuits, which may possess only resistance, or may be tuned to any frequency whatever. In many instances the object of employing the rejector circuit is to obtain maximum P.D. across its ends at a certain frequency (to which the rejector is tuned) while the P.D. set up by currents of other frequencies is required to be a minimum. Such a circuit is shown in fig. 44 in which e_r represents a source of alternating E.M.F. of frequency f_r and e_n a source of alternating E.M.F. having an amplitude equal to that of e_r but with a frequency variable from 0 to ∞ , the R.M.S. values being E_r , E_n . In series with the rejector circuit is a fixed resistance r , and it is desired to obtain as large a P.D. as possible

CHAPTER V.—PARA. 64

between the points A and B, at the frequency f_r , to which the rejector is tuned, while at any other frequency f_n the P.D. between A and B is to be as low as possible. The selectivity of this combination may then be defined as the ratio

$$\frac{\text{P.D. between A and B at the frequency } f_r}{\text{P.D. between A and B at any other frequency}} = \frac{V_r}{V_n}$$

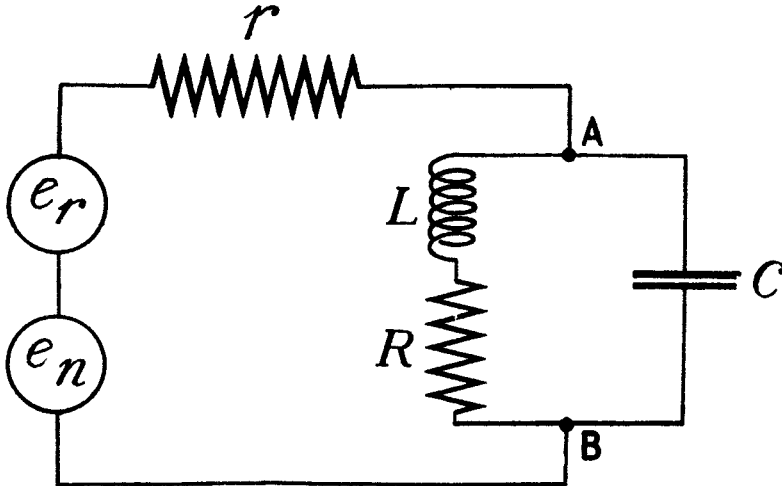


Fig. 44, CHAP. V.—Effect of external circuit upon selectivity of rejector circuit.

Referring to the diagram, the impedance operator, z , of the rejector circuit, at any frequency $\frac{\omega}{2\pi}$ is found by the rule for parallel impedances, which is identical with that for parallel resistances, provided the impedances are expressed in the form of vector operators. Denoting the impedance operator of the capacitive branch by z_c and that of the inductive branch by z_L , $z_c = \frac{1}{j\omega C}$ and $z_L = R + j\omega L$,

$$\begin{aligned} z &= \frac{z_c z_L}{z_c + z_L} \\ &= \frac{\frac{1}{j\omega C} (R + j\omega L)}{R + j\omega L + \frac{1}{j\omega C}} \\ &= \frac{R + j\omega L}{j\omega CR + 1 - \omega^2 LC} \end{aligned}$$

The current i , due to a sinusoidal E.M.F. e , is

$$i = \frac{e}{r + z}$$

and the P.D. between A and B is $v = z i$ or

$$v = \frac{z e}{r + z}$$

Denoting the impedance operator of the rejector at the frequency f_r by z_r and the impedance operator at any other frequency f_n by z_n

$$V_r = \frac{z_r \theta}{r + z_r}$$

$$V_n = \frac{z_n \theta}{r + z_n}$$

Since the amplitudes of e_r and e_n are numerically equal, however, the ratio $\frac{V_r}{V_n}$

$$\frac{V_r}{V_n} = \frac{z_r}{r + z_r} \div \frac{z_n}{r + z_n}$$

Inserting the values of z_r and z_n and observing that at the frequency f_r , $1 - \omega_r^2 LC = 0$,

$$\frac{V_r}{V_n} = \frac{\frac{R + j\omega_r L}{j\omega_r CR}}{r + \frac{R + j\omega_r L}{j\omega_r CR}} \times \frac{r + \frac{R + j\omega_n L}{j\omega_n CR + 1 - \omega_n^2 LC}}{\frac{R + j\omega_n L}{j\omega_n CR + 1 - \omega_n^2 LC}}$$

In the practical application of a circuit of this type R is always small compared with ωL and $R + j\omega L$ may be replaced without appreciable error by $j\omega L$.

$$\begin{aligned} \frac{V_r}{V_n} &= \frac{j\omega_r L}{j\omega_r (CRr + L)} \times \frac{r (j\omega_n CR + 1 - \omega_n^2 LC) + j\omega_n L}{j\omega_n L} \\ &= \frac{\omega_r}{\omega_n} \left\{ \frac{j\omega_n (CRr + L) + r (1 - \omega_n^2 LC)}{j\omega_r (CRr + L)} \right\} \end{aligned}$$

or, since $LC = \frac{1}{\omega_r^2}$

$$\begin{aligned} \frac{V_r}{V_n} &= \frac{\omega_r}{\omega_n} \left\{ \frac{j\omega_n (CRr + L) + r \left(1 - \frac{\omega_n^2}{\omega_r^2}\right)}{j\omega_r (CRr + L)} \right\} \\ &= \frac{\omega_r}{\omega_n} \left\{ \frac{\omega_n}{\omega_r} + \frac{r \left(1 - \frac{\omega_n^2}{\omega_r^2}\right)}{j\omega_r (CRr + L)} \right\} \\ &= 1 + \frac{r \left(1 - \frac{\omega_n^2}{\omega_r^2}\right)}{j \omega_n (CRr + L)} \end{aligned}$$

The second term of the right hand member of the equation may be further simplified :—

$$\frac{r \left(1 - \frac{\omega_n^2}{\omega_r^2}\right)}{j\omega_n (CRr + L)} = \frac{r (\omega_r^2 - \omega_n^2)}{j\omega_n \omega_r^3 (CRr + L)} = \frac{\omega_r^2 - \omega_n^2}{j\omega_n \left(\frac{R}{L} + \frac{1}{Cr}\right)}$$

hence

$$\frac{V_r}{V_n} = 1 - j \frac{\omega_r^2 - \omega_n^2}{\omega_n \left(\frac{R}{L} + \frac{1}{Cr}\right)}$$

CHAPTER V.—PARA. 64

In this form it may be seen that in order to obtain high selectivity the factor $\frac{R}{L} + \frac{1}{Cr}$ must be as small as possible. Thus in some practical applications of the rejector circuit selectivity is not necessarily achieved by making $\frac{L}{R}$ large, for the product LC is fixed by the value of the "desired" frequency that is by ω_r , hence a large value of L entails a small value of C , and consequently a large value of $\frac{1}{Cr}$. The effect will now be illustrated by a numerical example.

Example 13.—If $L = 160 \mu H$, $C = .00025 \mu F$, $R = 50$ ohms, $r = 5,000$ ohms, find the selectivity when $\omega_n = 5.6 \times 10^4$.

Considering the imaginary portion of (ii) only

$$\begin{aligned}\omega_r &= \sqrt{\frac{1}{LC}} = \sqrt{\frac{10^6}{160} \times \frac{10^6}{.00025}} = 5 \times 10^4 \\ \omega_r^2 - \omega_n^2 &= (\omega_r + \omega_n)(\omega_r - \omega_n) = 10.6 \times 10^4 \times .6 \times 10^4 \\ &= 6.36 \times 10^{12} \\ \frac{\omega_r^2 - \omega_n^2}{\omega_n} &= \frac{6.36 \times 10^{12}}{5.6 \times 10^4} = 1.135 \times 10^8 \\ \frac{R}{L} + \frac{1}{Cr} &= \frac{50}{160} \times 10^6 + \frac{10^6}{.00025 \times 5,000} \\ &= .3125 \times 10^6 + .8 \times 10^6 \\ &= 1.1125 \times 10^6 \\ \frac{V_r}{V_n} &= 1 - j \frac{1.135 \times 10^8}{1.1125 \times 10^6} \\ &= 1 - j 1.011.\end{aligned}$$

In R.M.S. values, therefore

$$\begin{aligned}V_r &= \sqrt{1^2 + (1.011)^2} V_n \\ \text{or } \frac{V_r}{V_n} &\doteq 1.42.\end{aligned}$$

Now suppose that r is increased to 100,000 ohms. Then $\frac{1}{Cr}$ becomes 4×10^4 which is small compared with $\frac{R}{L}$, and the latter factor will control the selectivity, giving

$$\begin{aligned}S &= 1 - j \frac{1.135 \times 10^8}{(.3125 + .04) \times 10^6} \\ &= 1 - j \frac{1.135}{.3525} \\ &= 1 - j 3.22\end{aligned}$$

That is, $V_r = \sqrt{1 + (3.22)^2} V_n$

$$\text{or } \frac{V_r}{V_n} = 3.375.$$

The voltage between A and B (fig. 44) due to the non-resonant frequency is in this instance only 29.6 per cent. of the voltage at resonance, for equal applied voltages.

If the value of the ratio $\frac{L}{C}$ is chosen so that it is equal to Rr or in other words if $\frac{L}{CR} = r$

the factor $\left(\frac{R}{L} + \frac{1}{Cr}\right)$ has its minimum value. This may be proved mathematically or by

taking simple numerical values. It follows that if in fig. 44 r represents the total internal resistance of a circuit which is equivalent in its effect to the generators e_r , e_n , specified above, the greatest selectivity will be obtained when the dynamic resistance of the rejector circuit is equal to the internal resistance of the equivalent generator. This is of the utmost importance in the design of the circuits used in connection with thermionic valves both for reception and transmission, and the above example shows the danger of making any assumption as to the selectivity of a rejector circuit without taking into account the circuits to which the rejector is connected.

CHAPTER VI.—THE TRANSFORMER

COUPLED RESONANT CIRCUITS

THE TRANSFORMER

1. One of the greatest advantages of an alternating current supply is the ease with which the transformation of a low to a high voltage or *vice versa* can be performed, without the aid of rotating machinery. This transformation is effected by means of a piece of apparatus called a static transformer, or more commonly, a transformer, which consists essentially of three circuits, namely, a magnetic circuit linking two electric circuits. The latter are termed the primary circuit, which contains the fundamental source of all the energy transformed or dissipated in the apparatus, and the secondary circuit respectively. The magnetic circuit is the volume of space occupied by magnetic flux, and may consist of a path of ferromagnetic material such as iron, or a non-magnetic material such as air, giving rise to what are called iron-core and air-core transformers respectively. The principles are the same in each type, but it is convenient to treat them in an entirely different manner, and in this chapter the term transformer may be regarded

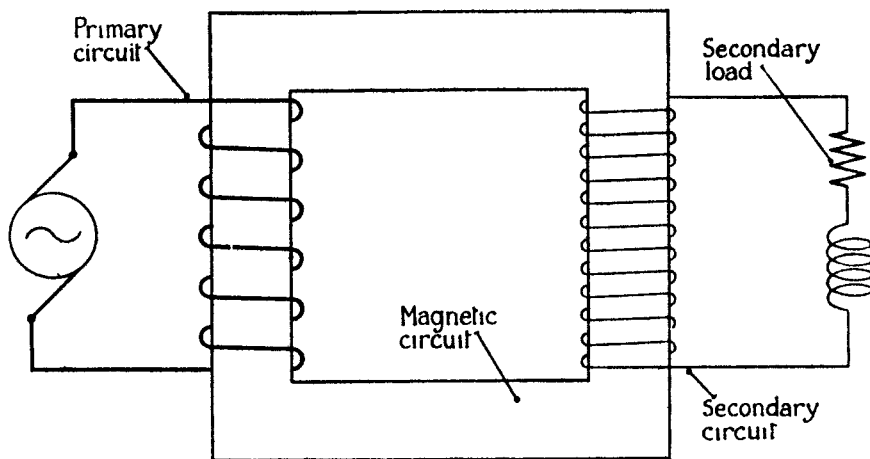


FIG. 1, CHAP. VI.—Principle of transformer.

as an abbreviation for iron-core transformer; the theory of air-core transformers is more conveniently treated in the section on "coupled circuits." The transformer is represented diagrammatically in fig. 1, and consists of a ferromagnetic core, which is built up of laminations, the best Swedish soft iron, or stalloy, being usually employed in order to reduce hysteresis and eddy current losses. These laminations are insulated from each other by varnish or thin paper. Over the core are wound the two electrically conductive circuits, primary and secondary, the windings being brought to terminals. If the secondary winding is given more turns than the primary, the transformer will give a secondary E.M.F. which is higher than the primary terminal P.D. and the contrary is the case if the secondary winding is given fewer turns than the primary. The two forms are therefore known as step-up and step-down transformers respectively. The two windings are well insulated from each other by micanite and varnished cloth, and if very high voltages are to be handled, possibly wound upon insulating sleeves. The general arrangement is shown in fig. 2, in which two principal types of construction are illustrated, namely the core type (a) and the shell type (b). It will be observed that the core type follows closely the theoretical diagram given previously, except that instead of carrying one winding on each of the two "limbs" of the core, both primary and secondary windings are distributed equally between the two. It may be said that in the core type the magnetic circuit is to a great extent covered by the coils, which are freely exposed, while in the shell type, the winding is chiefly enclosed by the magnetic circuit, which is itself exposed. It is usual to divide the high voltage winding into separately insulated sections in order to reduce the potential difference between the successive layers of wire. This leads to a decrease

CHAPTER VI.—PARA. 1

in the amount of insulation required by the wires themselves, the result being a net gain in cheapness and efficiency. Generally, the core is built in two or more parts so that the coils can be wound separately and afterwards placed in position, the portions of the core being after-

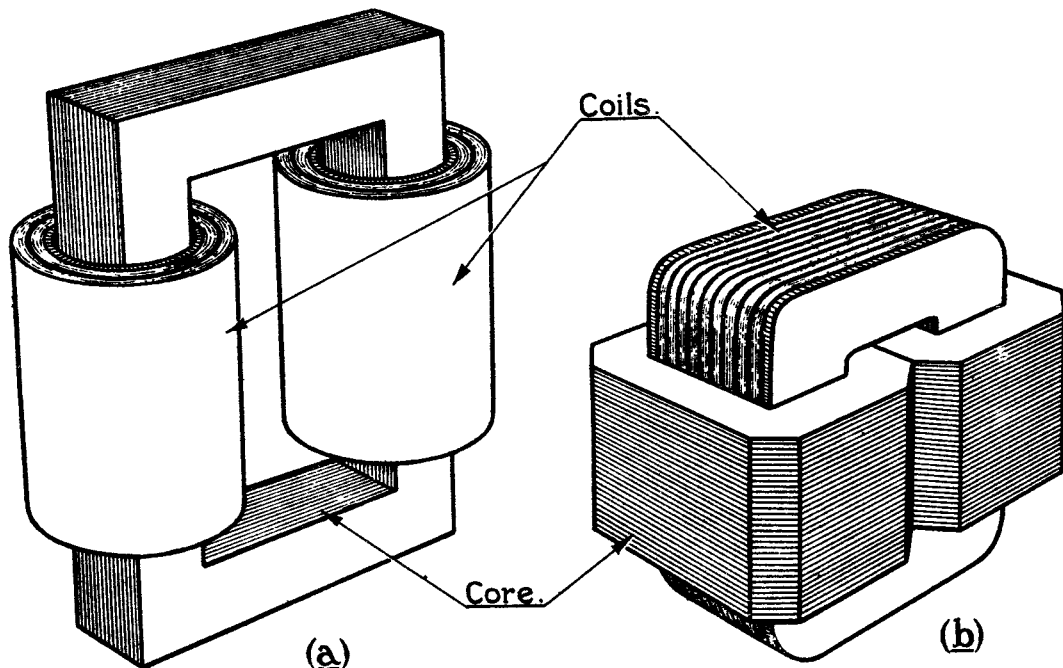


FIG. 2, CHAP. VI.—Transformers, core and shell types.

wards either butted together or preferably interleaved forming what is called an imbricated joint. Fig. 3 shews the latter type of core.

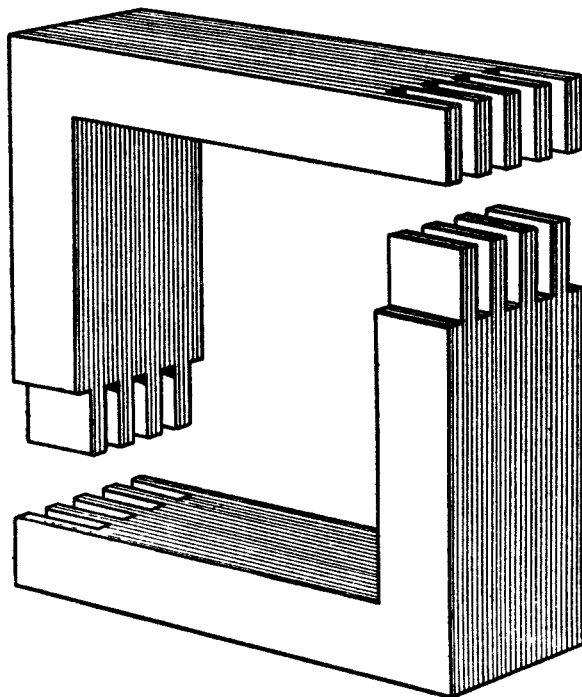


FIG. 3, CHAP. VI.—Imbricated joint in core.

Notation

2. Owing to the large number of variable quantities which must be taken into consideration when dealing with the action of the transformer, the vector diagrams are apt to appear rather complicated, but provided that the diagram is drawn step by step as detailed in the text, it will be found that the vectors themselves convey the whole of the essential theory.

The notation which is used is as follows :—

Primary terminal P.D.	V_p
Primary induced counter-E.M.F.	E_1
Component of terminal P.D. equal to latter	V_1
Secondary induced E.M.F.	E_2
Secondary terminal P.D.	V_2
Primary no-load current	I_0
Iron-loss component of I_0	I_i
Magnetising component of I_0	I_m
Primary load current	I_1
Secondary load current	I_2
Total primary current	I_p
Flux established by primary magnetising current	Φ_m
Flux due to primary load current	Φ_1
Flux due to secondary load current	Φ_2
Primary leakage flux	Φ_1
Secondary leakage flux	Φ_2
Counter E.M.F. due to primary leakage flux	E_1
Counter E.M.F. due to secondary leakage flux	E_2
Total fluxes and voltages allowing for magnetic leakage :—	
In primary circuit	Φ'_1, E'_1
In secondary circuit	Φ'_2, E'_2
Primary circuit constants :—	
Resistance of winding	R_1
Inductance of winding	L_1
Number of turns	N_1
Primary impedance, no load	Z_0
Primary inductance, no load	L_0
Effective reactance of primary, no load	X_0
Effective resistance of primary, no load	R_0
Equivalent primary resistance	R_p
Equivalent primary reactance	X_p
Secondary circuit constants :—	
Resistance of winding	R_2
Inductance of winding	L_2
Number of turns	N_2
Equivalent secondary resistance	R_s
Equivalent secondary reactance	X_s
Turns ratio, $\frac{N_2}{N_1}$	T

It will be assumed that the primary terminal P.D. is of sinusoidal waveform, and it will then only be necessary to deal with R.M.S. values of voltage, current and flux.

CHAPTER VI.—PARA. 3

Primary current and voltage, secondary on open circuit

3. If an alternator is connected to the primary terminals of the transformer, the secondary winding being left on open circuit, an alternating current will flow in the primary winding, and will establish a corresponding flux in the iron core. The primary winding is then acting simply as an inductive coil and the current will be equal to

$$I_o = \frac{V_p}{Z_o}$$

In all practical transformers the inductive reactance of the primary winding is very large indeed compared with its resistance, and I_o which is called the no-load primary current, will lag on the applied voltage by nearly 90° . The vector diagram fig. 4a shews the relative phases of I_o , the counter-E.M.F. of self-induction, E_1 , which is numerically equal to $\omega L_1 I_o$ and the reactive component, V_1 , of the applied voltage, which is equal and opposite to E_1 . A component $R_o I_o$ of the terminal voltage is also shewn; this provides for the unavoidable losses which occur even

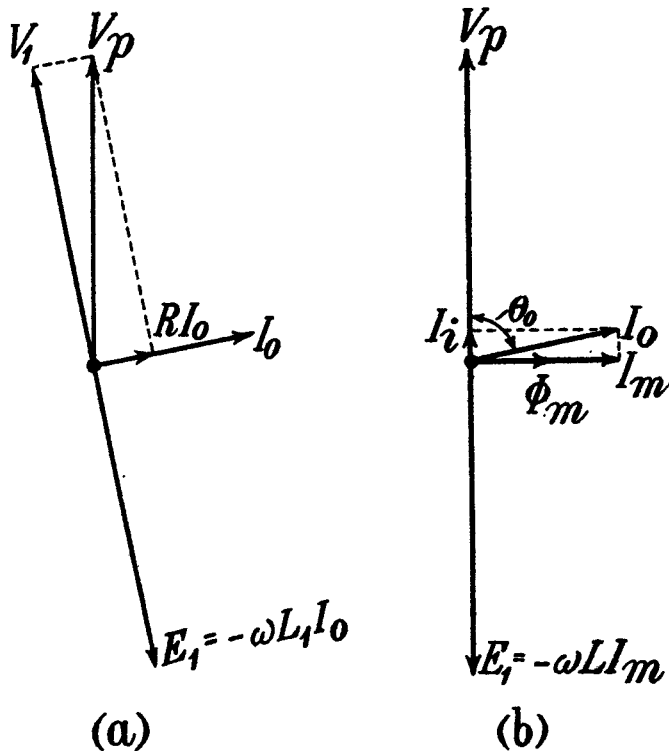


FIG. 4, CHAP. VI.—Vector diagram showing no-load conditions.

when the transformer secondary is on open circuit. This aspect of the no-load condition is identical with that taken in the discussion of an A.C. circuit possessing resistance and inductance in series (Chapter V). In dealing with the transformer, however, it is more convenient to consider that the resistance causing the no-load losses is in parallel with the primary winding. The no-load current then consists of two components in quadrature, first the loss component which is denoted by I_i and is in phase with the applied voltage V_p , and second a component I_m called the magnetising component which lags on the applied voltage by 90° , and is responsible for the establishment of the alternating flux Φ_m . The flux is in phase with the current I_m and the counter-E.M.F. of self-induction, E_1 , lags on I_m by 90° . The relationship between the various components of voltage and current is shewn in fig. 4b.

4. In the type of transformer under discussion, the primary and secondary windings are arranged on the core in such a way that practically all the flux produced by the magnetising current will link with every turn of the secondary winding; in consequence the flux will cause an induced E.M.F. not only in the primary but also in the secondary winding, the E.M.F. in every turn of winding being of the same magnitude. The total E.M.F. induced in the primary is by Faraday's law equal to N_1 times the rate of change of flux, or $N_1 \frac{d\Phi_m}{dt}$ and the secondary induced E.M.F. will be N_2 times the rate of change of flux or $N_2 \frac{d\Phi_m}{dt}$, N_1 and N_2 being the number of turns in the primary and secondary windings respectively. From this we may immediately deduce the most important law of the transformer.

$$\frac{\text{Secondary E.M.F.}}{\text{Primary counter-E.M.F.}} = \frac{\text{Turns on secondary winding}}{\text{Turns on primary winding}}$$

or, algebraically,

$$\frac{E_2}{E_1} = \frac{N_2}{N_1} = T$$

This is called the transformation ratio of the transformer. In a step-up transformer T is greater than unity, and in a step-down transformer, less than unity. As the secondary E.M.F. and the primary counter-E.M.F. are both caused by the same changing flux, they must be in phase with each other, and since the primary counter-E.M.F. is (practically) 180° out of phase with the voltage applied to the primary, the secondary induced E.M.F. must also be 180° out of phase with the applied voltage. To illustrate the relation between the voltage applied to the primary and the induced secondary E.M.F. suppose a transformer to have 100 primary turns and 3,300 secondary turns the secondary being on open circuit. An applied voltage of 220 volts will cause a small current to flow, and the counter-E.M.F. induced in the primary winding will be nearly but not quite 220 volts, the difference being utilised to supply the small power losses. The counter-E.M.F. induced in each turn of primary winding will be $\frac{220}{100} = 2.2$ volts per turn, and a similar E.M.F. will also be induced in each turn of secondary winding, hence the total secondary E.M.F. will be 2.2×3300 or 7260 volts.

Effect of secondary load current

5. (i) If a circuit is connected to the secondary terminals, the induced E.M.F. in the latter winding will establish a secondary load current. This current in its turn will set up in the core a magnetic flux which is proportional to the ampere-turns on the secondary winding. The establishment of a secondary current therefore tends to weaken the flux caused by the primary current, but actually this effect does not occur, because as soon as secondary current starts to flow, reducing the total effective flux, the primary counter-E.M.F. is also reduced, and the applied voltage is able to supply an increased primary current called the primary load current, which in its turn sets up a flux which is equal and opposite to that caused by the secondary load current. This equality must signify that the additional primary ampere-turns are equal to the secondary ampere-turns or Primary load current $\times N_1 =$ Secondary load current $\times N_2$.

$$\frac{\text{Primary load current}}{\text{Secondary load current}} = \frac{N_2}{N_1} = T.$$

If I_1 and I_2 are the primary and secondary load currents

$$I_1 = TI_2$$

The total primary current is the vector sum of the original no-load current and the primary load current. The magnitude of the latter will increase directly with the magnitude of the

CHAPTER VI.—PARA. 6

secondary current, and when the secondary is supplying a current approaching the normal load for which it is designed, the primary load current is so much greater than the no-load component that the latter may be disregarded. Under these conditions, the relation between primary and secondary currents may be written

$$I_p = TI_2$$

where I_p is the total primary current.

(ii) The phase relation between the primary current and terminal P.D. depends upon the nature of the impedance of the appliance to which the secondary winding is supplying current and can be shewn by vector diagrams. In order to preserve balance in these, the following convention is adopted. If the transformation ratio is less than unity (step-down) the vectors representing secondary voltages are drawn to a scale T times as large as the primary voltages,

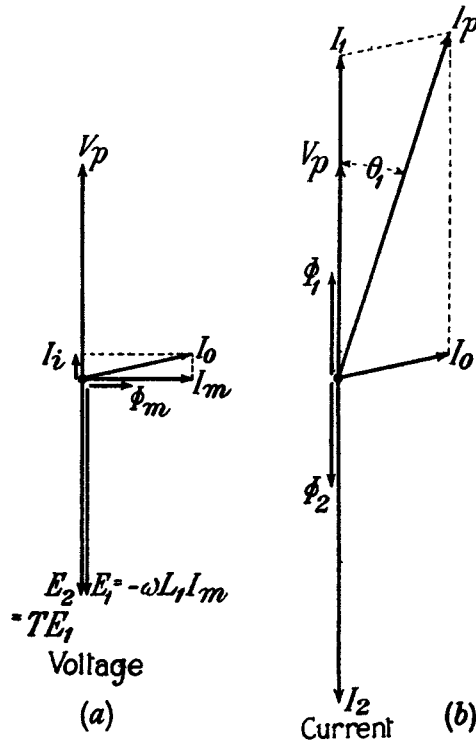


FIG. 5, CHAP. VI.—Voltage and current relations with purely resistive load.

while the primary currents are drawn on a scale T times as large as the secondary currents. If the transformation ratio is greater than unity these scales are adjusted in an opposite manner. In effect therefore the vector diagrams are drawn as for a transformer of unity turns ratio.

Resistive load on secondary

6. The effect of connecting to the secondary winding a circuit possessing resistance only is shewn by the vector diagrams, fig. 5. In fig. 5a I_m is the magnetising component of the primary no-load current I_o , I_i the loss component. The primary flux Φ_m is caused by I_m and is in phase with the latter. Lagging upon this flux by 90° are the induced voltages E_1 and E_2 , the former being the primary counter-E.M.F. and the latter the secondary induced E.M.F. E_1 is to all intents and purposes equal to the applied voltage V_p and is 180° out of phase with it, while E_2 is in phase with E_1 , hence the secondary E.M.F. is 180° out of phase with the applied voltage.

Now consider the effects caused by a secondary load current I_2 which is in phase with the secondary E.M.F., E_2 . This current will produce a flux Φ_2 in phase with I_2 as shewn in fig. 5b, and as already stated the effect of this flux is to reduce the total flux linking with the primary winding, so that the inductance of the latter is decreased. As a result an increased primary current will flow setting up a flux Φ_1 which is equal and opposite to Φ_2 . The flux Φ_1 is caused by a component of primary current which is in phase with it and is termed the primary load current I_1 , and it is apparent that I_1 must be exactly opposite in phase to I_2 in order that the phase relationship between Φ_2 and Φ_1 shall be correct. The total primary current I_p is the vector sum of the primary load current I_1 and the no-load current I_m , and is obtained by the usual construction. It will be seen that I_p lags upon the applied voltage V_p by an angle θ_1 which is inversely proportional to the load current, i.e. with a small secondary current, θ_1 is nearly 90° but when the secondary winding is caused to give a large current the primary current and primary voltage are very nearly in phase. This may be expressed by stating that when on

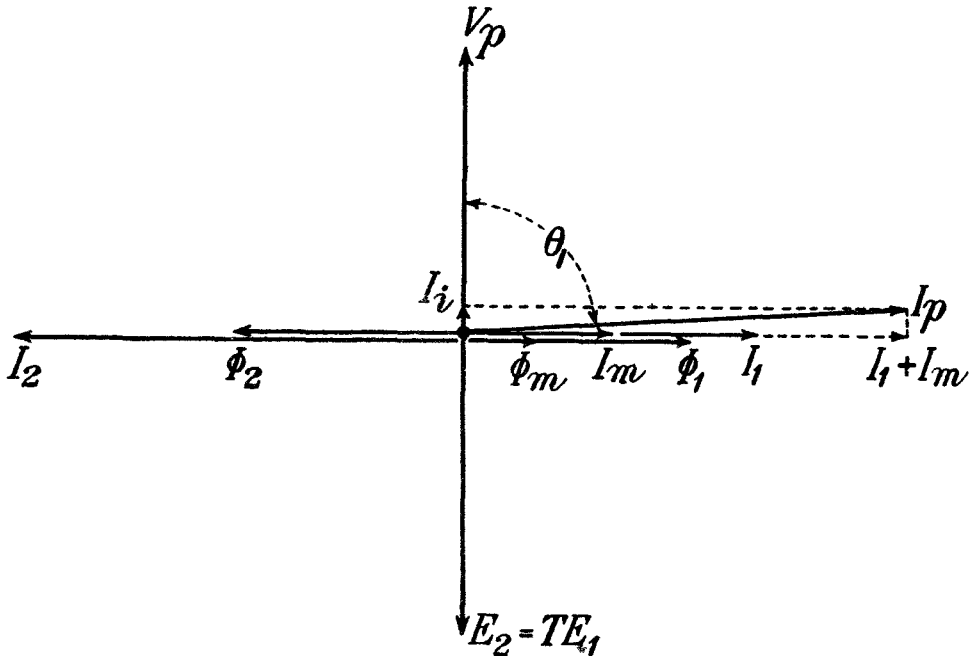


FIG. 6, CHAP. VI.—Voltage and current relations with purely inductive load.

full load with a purely resistive circuit the transformer acts as if it were practically noninductive, or alternatively, that with such conditions of loading the power factor of the primary circuit approaches unity.

Purely reactive loading

7. (i) If the secondary load is replaced by an inductance having negligible resistance the resulting conditions are shewn in the vector diagram fig. 6. Here again I_m is the magnetising current and the flux Φ_m is in phase with it. The vectors E_1 and E_2 are the primary and secondary induced voltages as before and V_p equal in magnitude to E_1 , but of opposite phase, is the primary terminal voltage. It should be noted that in this and all future vector diagrams, the vector E_1 is not actually drawn, because it coincides with E_2 owing to the adjustment of relative scales of current and voltage, paragraph 5. The secondary current I_2 will lag on the secondary voltage E_2 by 90° , and will set up the flux Φ_2 , but in order that the total flux may remain constant and equal to Φ_m , a primary load current I_1 will be established, its phase being such that its resultant flux

CHAPTER VI.—PARA. 8

Φ_1 is equal and opposite to Φ_2 . The primary load current lags on the applied voltage by 90° , and the total primary current I_p which is the vector sum of I_1 , I_m and I_i , will lag on V_p by an angle θ_1 which is nearly but not quite 90° .

(ii) The effect of a purely capacitive secondary load is shown in the vector diagram, fig. 7, the magnetising current I_m and flux Φ_m being as before, and E_2 the secondary E.M.F. lags by 90° on these. The secondary load current I_2 leads on E_2 by 90° , and the primary load current I_1 is 180° out of phase with I_2 . The primary load current I_p is the vector sum of I_1 and I_o . Under certain conditions, a resonant condition may exist, and this occurs when the primary load current is exactly equal (and of opposite phase) to the magnetising current. The total primary current I_p is then only that due to the iron losses, and so far as the primary terminal

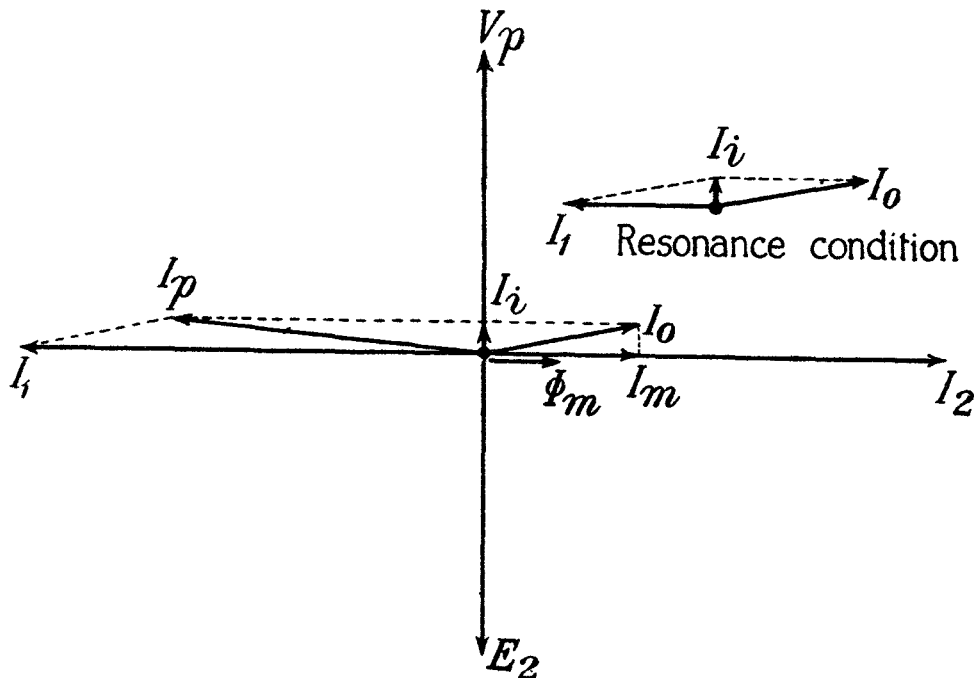


FIG. 7, CHAP. VI.—Voltage and current relations with purely capacitive load.

P.D. is concerned the transformer with its capacitance load behaves like a rejector circuit. This condition never occurs in power transformers but is of importance in the small iron-core transformer used in audio frequency amplifiers.

Secondary load both resistive and reactive

8. (i) As a rule, the appliance to which the secondary winding is connected will possess both reactance and resistance. In order to investigate the action under such conditions, consider the secondary load to consist of a resistance of r_2 ohms and an inductance of l_2 henries connected in series, and the impedance of the load will then be $\sqrt{r_2^2 + \omega^2 l_2^2} = Z_2$ ohms. The vector diagram is given in fig. 8, and as before, the secondary induced E.M.F. E_2 will cause a current I_2 to flow, its magnitude being $I_2 = \frac{E_2}{Z_2}$ ohms; this current will lag on by an angle $\theta_2 = \frac{\tan^{-1} \omega l_2}{r_2}$

The primary load current I_1 is of opposite phase to I_2 in order to establish the necessary counter-balancing flux, and the total primary current is the vector sum of I_1 and I_o , lagging upon the

applied voltage V_p by the angle θ_1 which is slightly greater than θ_2 , the inequality becoming smaller as the secondary current increases. When the transformer is giving its full secondary current, it may be considered that the primary current lags on the applied E.M.F. by an angle equal to θ_2 which is tantamount to an assertion that the transformer itself has no effect upon the phase difference, but merely serves to alter the voltage applied to the external circuit connected to the secondary terminals.

(ii) In fig. 9 is shewn the condition which arises when the secondary load possesses both resistance and capacitive reactance. Here E_1 and E_2 lag by 90° on the magnetising current, and I_2 leads on E_2 by an angle θ_2 , the primary load current I_1 being of exactly opposite phase in order to produce the necessary counter-balancing flux. The primary current I_p is the vector

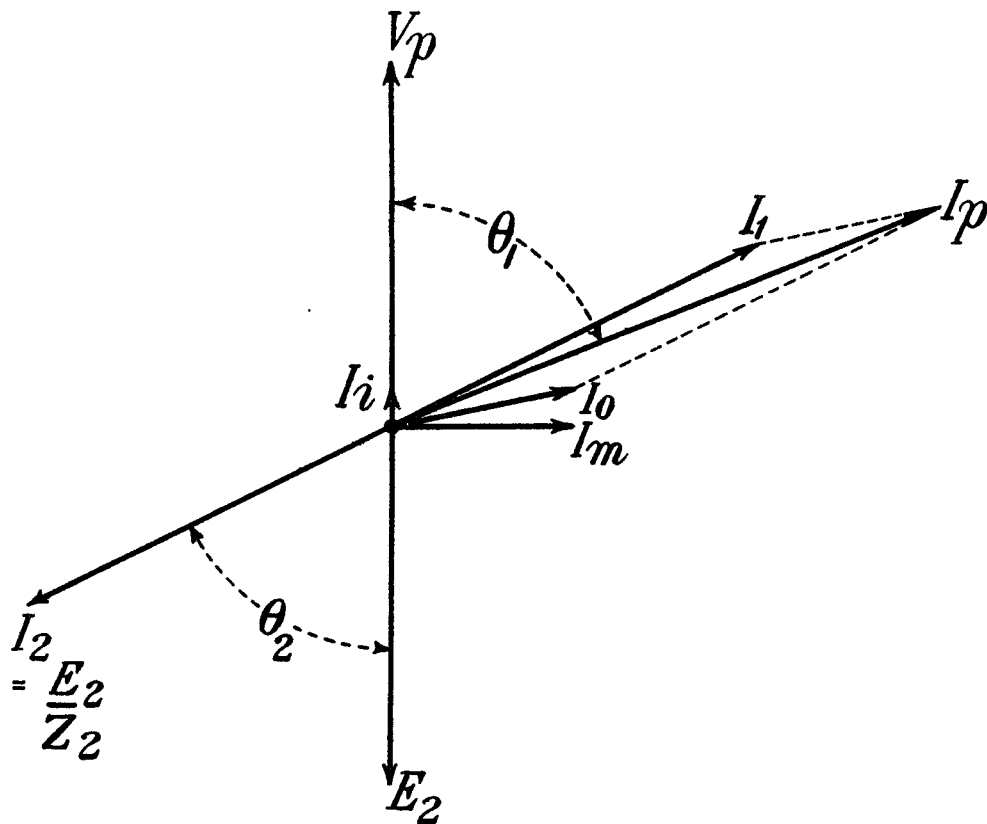


FIG. 8, CHAP. VI.—Voltage and current relations with load possessing both resistance and inductance.

sum of I_1 and I_o , and leads on the applied voltage V_p by the angle θ_1 which is slightly less than θ_2 . In the inset diagram the resonance condition is depicted. In this case the vector sum of I_1 and I_o , i.e. I_p is in phase with the applied voltage V_p , and the circuit behaves as if it possesses neither inductance nor capacitance, imposing only a resistive load upon the alternator.

Losses in transformers

9. In the circuits just discussed, it will be observed that, if the no load current is neglected, the following relations hold, namely $I_p = I_2 \times T$, $E_2 = E_1 \times T = V_p \times T$. From this it follows that $I_p \times V_p = I_2 \times T \times \frac{E_2}{T}$ or $I_p V_p = I_2 E_2$. This indicates that if the losses are

CHAPTER VI.—PARA. 10

neglected, the power input is equal to the power output, which is of course to be expected from the principle of conservation of energy. In practice this equality is not achieved owing to the conversion of a portion of the energy supplied into heat, and it is convenient to divide the losses according to the portion of the transformer in which they occur. This leads to the conception of (i) iron losses (ii) copper losses and (iii) losses due to magnetic leakage.

Iron losses

10. The iron losses are caused by (a) eddy currents in the iron, and (b) hysteresis.

Eddy currents.—These are due to the effects already noted in connection with the armatures of dynamo-electric machinery (Chapter IV), and are reduced but not entirely eliminated by lamination of the iron core. The degree to which the sub-division is performed depends to some

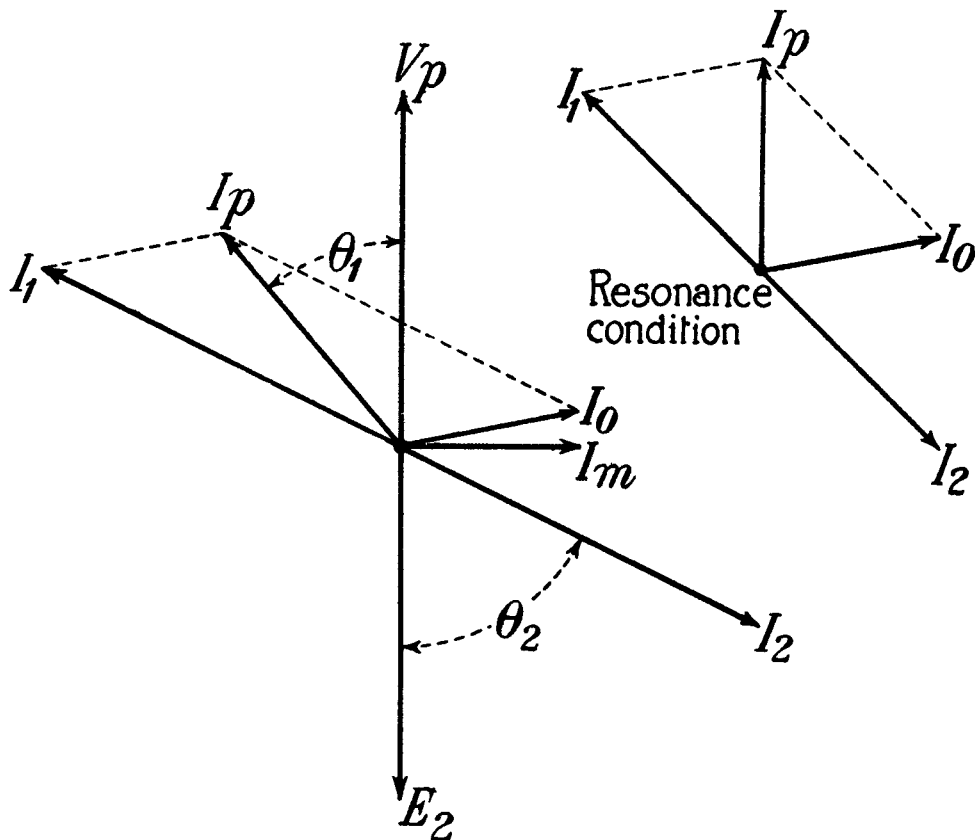


FIG. 9, CHAP. VI.—Voltage and current relations with load possessing both resistance and capacitance.

extent upon the kind of iron employed, and also upon the frequency for which the transformer is designed. At the standard commercial frequency (50 cycles per second) a thickness of $\cdot 014$ inch is frequently adopted if soft iron is used, but with stalloy the laminations may be as much as $\cdot 03$ inch in thickness. At lower frequencies thicker laminations may be used, but at higher frequencies it becomes increasingly difficult and expensive to achieve the degree of lamination which is necessary if the iron losses are not to be increased. At frequencies of the order of hundreds or thousands per second therefore, the iron losses are much greater than at commercial frequencies.

Hysteresis losses.—During each alteration of the magnetising current, the core is carried through a complete cycle of magnetisation, and the variation of the flux density B under the influence of the magnetising force H gives rise to a B/H curve which is actually a closed loop. This signifies that the energy expended in creating the flux is not entirely restored to the electric circuit when the flux is destroyed, or that energy is expended in changing the magnetic state of the material. This phenomenon was dealt with in Chapter II, and it is only necessary to add that in practice it is usual to assume that the energy expended may be represented by an equation of the form

$$W_h = hB^{1.7}$$

where W_h is the energy required to carry the iron through a complete magnetic cycle, in ergs per cubic centimetre, and h a constant for the particular kind of iron. For materials customarily employed in transformers h usually lies between .001 and .0015. As an example, consider the power P_h required to overcome the hysteresis loss in a transformer core, the cross section of which is 100 square centimetres and the length of the mean magnetic circuit is 200 centimetres. The hysteresis constant h may be taken as .001, and the frequency 50 cycles per second, while the flux density may be as high as 10,000 C.G.S. units.

$$\begin{aligned} \text{Then } P_h &= .001 \times 10000^{1.7} \times 10^{-7} \text{ joules per cm}^3 \text{ per period} \\ &= .001 \times 10000^{1.7} \times 10^{-7} \times 100 \times 200 \times 50 \frac{\text{joules}}{\text{sec.}} \quad (10000^{1.7} = 6,310,000) \\ \therefore P_h &= .001 \times 6.31 \times 10^6 \times 10^{-7} \times 10^2 \times 2 \times 10^2 \times 50 \\ &= 10^3 \times 6.31 \times 10^6 \times 10^{-7} \times 10^2 \times 10^2 \times 10^2 = 631 \text{ watts.} \end{aligned}$$

The iron losses are practically the same on no load as on full load, and in the former condition are responsible for practically the whole of the losses, because the actual ohmic resistance of the primary is so small that, when only the small no-load primary current is flowing, the I^2R loss in the conductor is negligible. It will be remembered that the no load current was divided into two components, one of which, I_m , was said to establish the flux and to cause no loss of energy, while the other, I_h , was held responsible for the losses under no-load conditions. As we have now seen, these are almost entirely due to the iron and consequently I_i is referred to as the iron-loss current.

Copper losses

11. These are caused by the actual ohmic resistance of the windings, and as stated above are negligible when the secondary is unloaded. When a secondary current is established, however, the primary current increases and the copper losses in the primary circuit are equal to $I_1^2 R_1$, while the copper losses in the secondary circuit become $I_2^2 R_2$, hence the total copper losses are

$$P_c = I_1^2 R_1 + I_2^2 R_2 \text{ watts.}$$

In consequence of this waste of power the resistance is responsible for a voltage drop in each winding, and to produce a given number of magnetising ampereturns, the primary terminal voltage must be greater than that calculated by ignoring the effect of resistance, while the resistance of the secondary winding causes its terminal P.D. to be less than the induced E.M.F. E_2 . These consequences are shown in a vector diagram (fig. 10). The no-load current I_0 and the voltages E_1 and E_2 are as in previous diagrams. Assuming that the load possesses both inductance and resistance, the secondary current I_2 lags upon E_2 by the angle θ_2 , and a component of voltage $e_2 = I_2 R_2$ in phase with I_2 represents the internal secondary P.D. due to the copper loss. The secondary terminal voltage is the vector difference between E_2 and e_2 and is shown by the vector V_2 . Similar effects occur in the primary, and it is seen that I_1 , the primary current is the vector sum of I_0 and I_1 the no-load and load currents respectively. The copper loss in this circuit is responsible for an internal P.D. equal to $I_1 R_1 = e_1$ which is in phase with I_1 , and the total voltage V_p which must be applied to the primary is now not V_1 equal and opposite to E_1 , as in previous instances, but the vector sum of e_1 and V_1 .

CHAPTER VI.—PARA. 11

It has been shown that the total copper losses are equal to $I_p^2 R_1 + I_s^2 R_2$ watts. It is sometimes convenient to regard these as being caused by a single resistance placed either in the primary or secondary circuit. As I_2 is practically equal to $\frac{I_p}{T}$, the copper losses may be written $I_p^2 R_1 + I_p^2 \frac{R_2}{T^2}$ or $I_p^2 \left(R_1 + \frac{R_2}{T^2} \right)$ watts and if no copper losses existed in the transformer itself a resistance $R_1 + \frac{R_2}{T^2}$ placed in series with the primary circuit would have the same effect as the

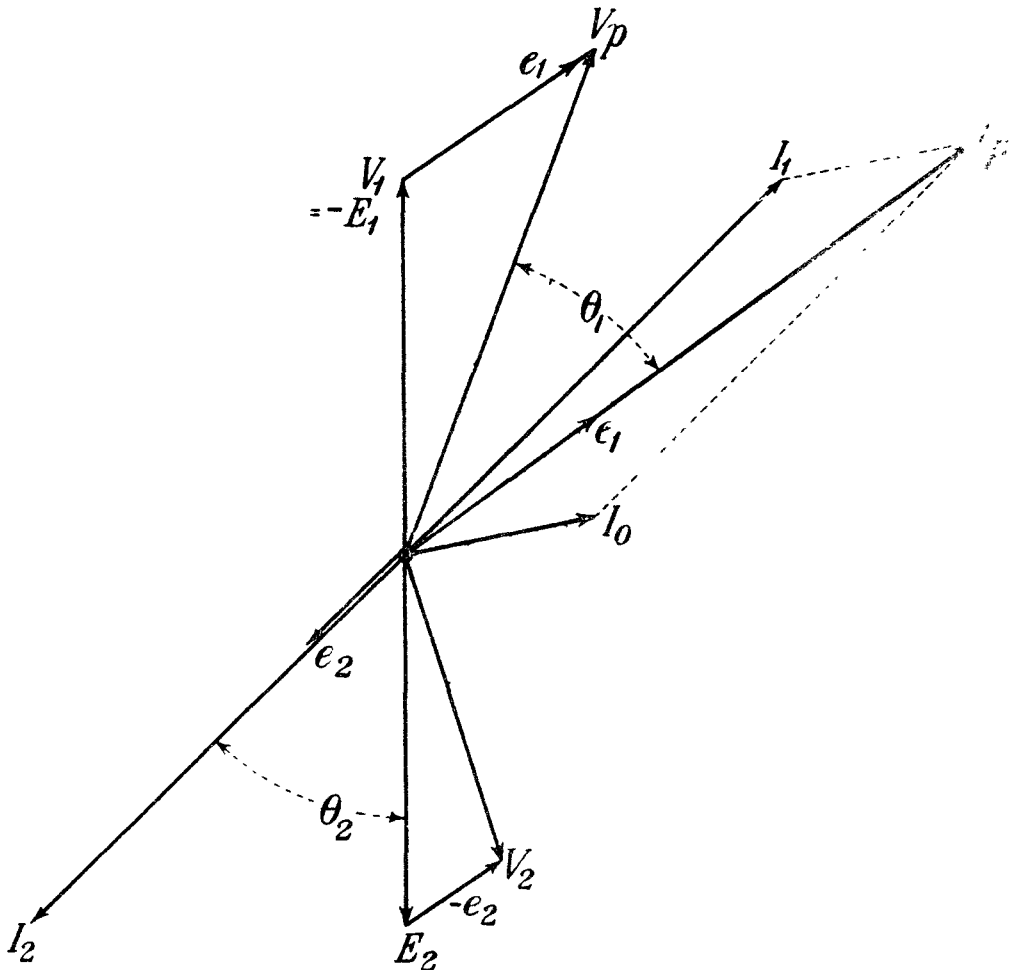


FIG. 10, CHAP. VI.—Effect of internal resistance of windings.

actual primary and secondary resistances, both with respect to the losses and to the P.D. at the secondary terminals. The expression $\left(R_1 + \frac{R_2}{T^2} \right)$ is therefore called the equivalent resistance of the transformer referred to the primary circuit, or more briefly as the equivalent primary resistance and is generally denoted by R_p . Alternatively, it may be assumed that the whole of the

copper losses are caused by a resistance equal to $R_2 + T^2 R_1$ in the secondary circuit, the primary circuit then being assumed to have no copper losses; the expression $(R_2 + T^2 R_1)$ is called the equivalent secondary resistance, and is denoted by R_s . It may be observed that if the windings are so designed that the current density is the same in both primary and secondary windings, that is, if each winding carries the same current per square millimetre of cross-sectional area, the cross section of each winding is inversely proportional to the number of turns, while the length of each winding is proportional to the number of turns, provided that the primary and secondary are intermingled and not wound one over the other. In consequence, $T^2 R_1$ is always approximately equal to R_2 , and the equivalent resistance approximately equal to double the true resistance, i.e. $R_p \doteq 2R_1$, or $R_s \doteq 2R_2$.

Example.—A 25 K.V.A. transformer has a step up of 10 to 1, and is designed for a primary voltage of 250 volts. The primary winding has a resistance of $\cdot 02$ ohms and the secondary $1\cdot 8$ ohms. Find the equivalent secondary resistance, the fall of terminal P.D. at full load, and the copper losses at full load.

$$\begin{aligned}\text{The equivalent secondary resistance} &= R_2 + T^2 R_1 \\ &= 1\cdot 8 + 10^2 \times \cdot 02 \\ &= 3\cdot 8 \text{ ohms.}\end{aligned}$$

$$\text{N.B.}—2R_2 = 3\cdot 6 \text{ ohms.}$$

Full load primary current (neglecting no-load component)

$$= \frac{25000 \text{ volt-amperes}}{250 \text{ volts}} = 100 \text{ amperes.}$$

$$\begin{aligned}\text{The voltage drop due to primary resistance} &= 100 \times \cdot 02 \\ &= 2 \text{ volts.}\end{aligned}$$

Hence the secondary voltage, instead of being 2500 volts will be 2480 volts.

$$\text{Full load secondary current} = \frac{I_p}{T} = \frac{100}{10} = 10 \text{ amperes.}$$

$$\begin{aligned}\text{Voltage drop due to secondary resistance} &= 10 \times 1\cdot 8 = 18 \text{ volts.} \\ \text{Hence the secondary terminal P.D.} &= 2480 - 18 \\ &= 2462 \text{ volts.}\end{aligned}$$

$$\begin{aligned}\text{Copper loss at full load} &= I_p^2 R_1 + I_s^2 R_2 \\ &= 100^2 \times \cdot 02 + 10^2 \times 1\cdot 8 \\ &= 200 + 180 \\ &= 380 \text{ watts.}\end{aligned}$$

Using the equivalent secondary resistance :—

$$\text{Total resistance drop} = I_2 (R_2 + T^2 R_1) = I_2 R_s = 10 \times 3\cdot 8 = 38 \text{ volts.}$$

$$\text{Secondary terminal P.D.} = 2500 - 38 = 2462 \text{ volts.}$$

$$\begin{aligned}\text{Copper losses} &= I_s^2 (R_2 + T^2 R_1) \\ &= 100 \times 3\cdot 8 = 380 \text{ watts.}\end{aligned}$$

The use of the “equivalent resistance” method obviously reduces the amount of arithmetic required.

Losses due to magnetic leakage

12. Up to the present it has been assumed that the whole of the flux set up by the magnetising current passes through every turn on both windings of the transformer, but this is not entirely true. Considering a transformer with the general design shown in fig. 11; the current in the primary winding sets up a flux, a portion of which may leave the core and pass through the air, and these tubes of flux do not link with the secondary winding. As they do not assist in producing a secondary E.M.F. they are known collectively as the primary leakage flux. Similarly,

CHAPTER VI.—PARA. 13

the current flowing in the secondary will produce some tubes of flux which do not link with the primary winding, and these constitute the secondary leakage flux. The leakage flux due to each winding is produced by the current in that winding and is independent of the other, and each portion is therefore in phase with and proportional to the current responsible for its existence, both leakage fluxes increasing with an increase of load current. This is in contrast with the main flux linking with both windings, which being proportional to the magnetising current I_m remains constant at all loads. Since the primary ampereturns and secondary ampereturns are nearly equal the leakage fluxes will also approach equality, because the reluctance of their

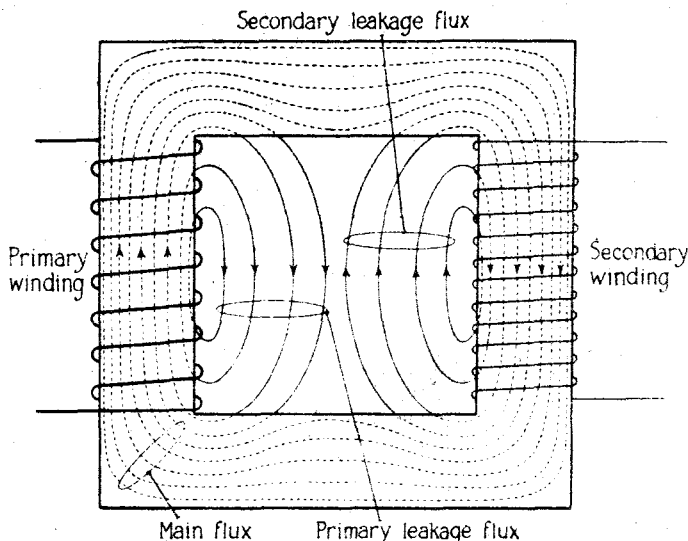


FIG. 11, CHAP. VI.—Primary and secondary leakage flux.

respective paths cannot be very different. The effects of magnetic leakage are shewn in the vector diagram fig. 12, in which the common flux is shewn by Φ_m , and lagging on this by 90° are the induced voltages E_1 and E_2 in the primary and secondary windings respectively. E_1 is counterbalanced by a component of the applied voltage which is represented by V_1 . The secondary voltage E_2 gives rise to the secondary current I_2 , and I_2 in turn produces the flux Φ_2 of which a portion Φ_{l2} represents the leakage flux. The actual secondary flux is the vector sum of the common flux Φ_m and the secondary leakage flux Φ_{l2} , and is indicated by the vector Φ'_2 . The effect of the secondary leakage flux is to cause a reduction in the total induced secondary voltage, which may be considered to be due to a component of the secondary E.M.F. which lags by 90° on the leakage flux, as shewn by the vector $\mathcal{L}E_2$. The effective E.M.F. acting in the secondary winding is the sum of the vectors E_2 and $\mathcal{L}E_2$ and is given by E'_2 . It will be observed that this vector lags by 90° on the effective secondary flux Φ'_2 . The primary load current I_1 produces the flux Φ_1 counterbalancing the flux Φ_2 , and the primary current I_p is the vector sum of I_1 and the no-load current I_0 . The primary leakage flux Φ_{l1} is in phase with I_p , and the total primary flux is therefore Φ'_1 . The result of the leakage flux may be represented by considering that an additional counter E.M.F. $\mathcal{L}E_1$ is induced in the primary winding, and this must be overcome by an additional applied voltage, hence the terminal P.D. is not V_1 , 180° out of phase with E_2 , but V_p which leads upon the resultant primary flux Φ'_1 by 90° .

Effects of magnetic leakage

13. (i) It will be observed that owing to magnetic leakage, the applied E.M.F. has to be increased in value in order to obtain a given secondary current, and this applied E.M.F. leads on the primary current by a greater angle than would be the case in the absence of leakage. The latter therefore appears to add an effective inductance in series with the primary winding, and

this is called the leakage inductance of the primary. Again in the secondary circuit the terminal voltage is reduced by the effect of leakage from E_2 lagging by 90° on the common flux, to E'_2 lagging by an angle greater than 90° . The secondary current I_2 lags on the actual secondary E.M.F. E'_2 by the angle θ_2 , whereas in the absence of the secondary leakage flux it would lag by the same angle upon the voltage E_2 . The effect of the magnetic leakage is therefore to increase the angle of lag between the secondary current and the common flux, and is equivalent to the

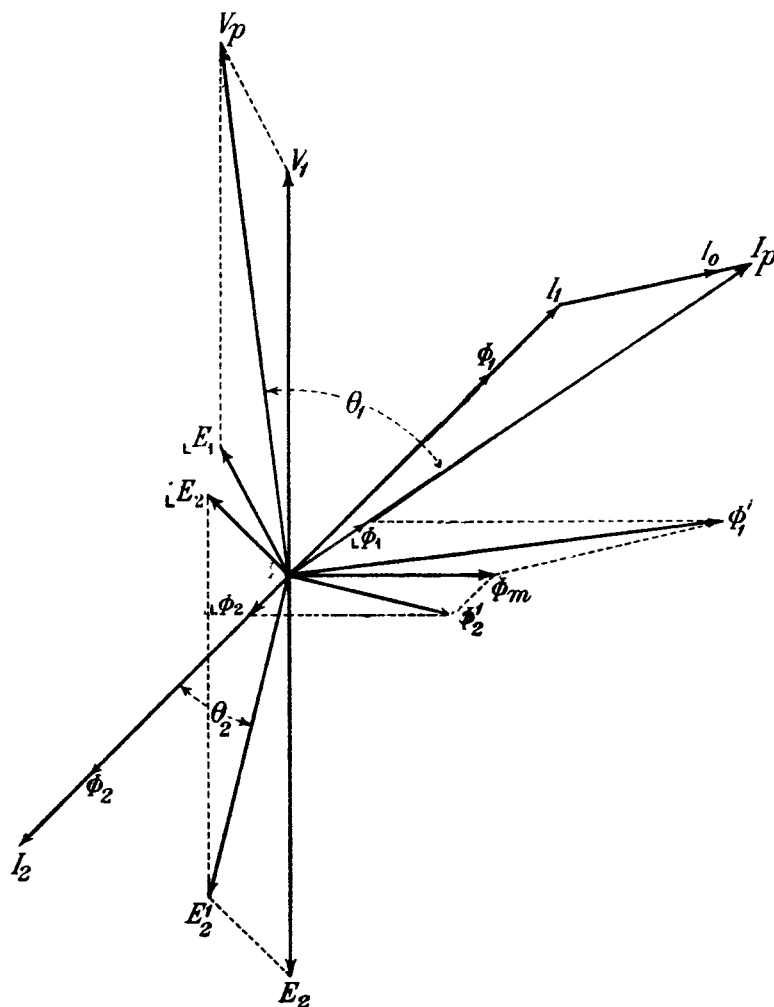


FIG. 12, CHAP. VI.—Effect of leakage flux.

introduction of an inductive reactance into the secondary circuit. Both leakage inductances cause a reduction in the secondary terminal P.D. for a given applied voltage, and in practice certain steps are taken to reduce this reactive voltage drop. One expedient, applicable chiefly to transformers of the core type, is to use concentric windings, the secondary being wound in sections over the primary. In the shell type of transformer, the windings may be sectionally wound and the primary and secondary sections intermingled upon the centre limb, with the same object. The arrangement of one winding on each limb of the core, as in fig. 1, is hardly ever adopted.

CHAPTER VI.—PARA. 13

(ii) The effects of both copper losses and magnetic leakage are illustrated by the vector diagram given in fig. 13, which shews the conditions when the secondary load is both resistive and inductive. The secondary current I_2 lags on the secondary induced E.M.F., and in phase with I_2 is the internal voltage drop I_2R_2 , while the reactive drop I_2X_2 due to magnetic leakage leads on I_2 by 90° . The total internal voltage drop is the vector sum of I_2R_2 and I_2X_2 and is given by I_2Z_2 , the secondary terminal P.D. being therefore V_2 , which is the result of subtracting

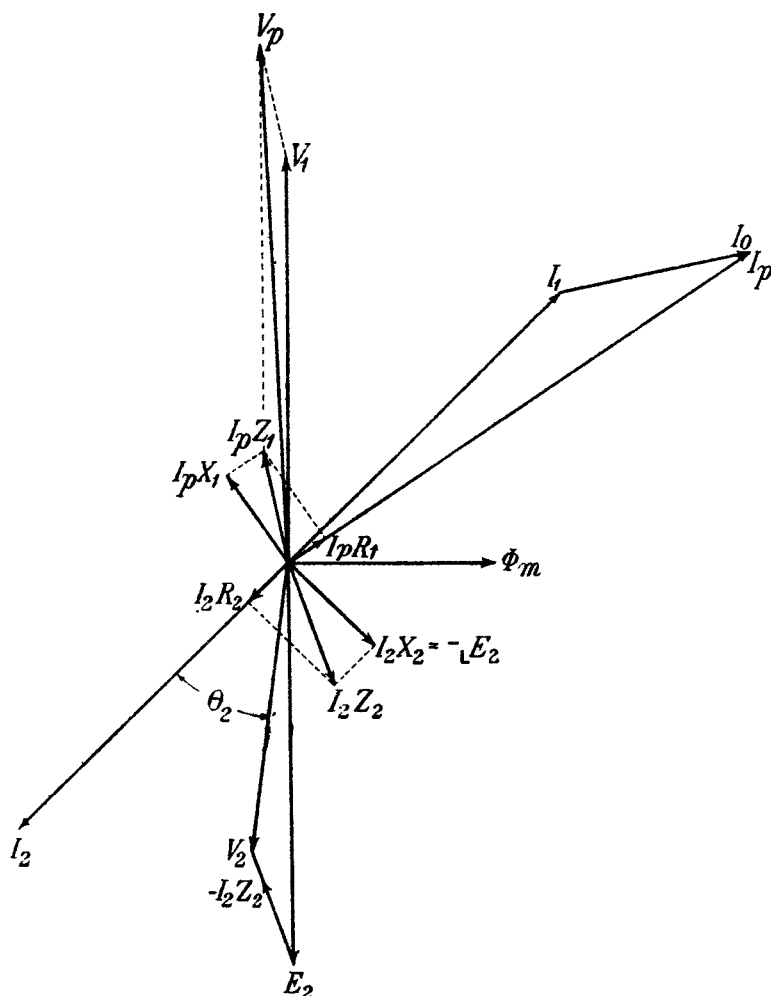


FIG. 13, CHAP. VI.—Effect of leakage flux and internal resistance.

the internal volts drop I_2Z_2 from the secondary E.M.F. E_2 . I_2 is seen to lag on V_2 by an angle θ_2 . The primary current I_p is the sum of the primary load current I_1 and the no-load current I_0 , and the internal resistive volts drop I_pR_1 is in phase with I_p while the reactive volts drop I_pX_1 leads on I_p by 90° . The total internal primary volts drop is the vector sum of I_pR_1 and I_pX_1 , and is shewn in the diagram by I_pZ_1 . The primary applied voltage V_p is the vector sum of the internal drop I_pZ_1 and the component V_1 which is equal and opposite to the primary counter-E.M.F.

Equivalent circuit of the transformer

14. The behaviour of a power transformer may be represented with a high degree of accuracy by an equivalent circuit, in which the transformer itself is considered to be an ideal one, giving a secondary E.M.F. equal to the primary terminal voltage multiplied by the turns ratio, and having no internal losses whatever. The primary resistance R_1 and leakage reactance X_1 are considered to exist outside the actual transformer, and the same applies to the secondary resistance R_2 and leakage reactance X_2 . As it must be assumed that no current will flow through the primary winding of this ideal transformer unless the secondary is on load, it is necessary to add an imaginary circuit which will account theoretically for the no-load current, and this is done by inserting a circuit consisting of a resistance R_0 and inductive reactance X_0 in parallel

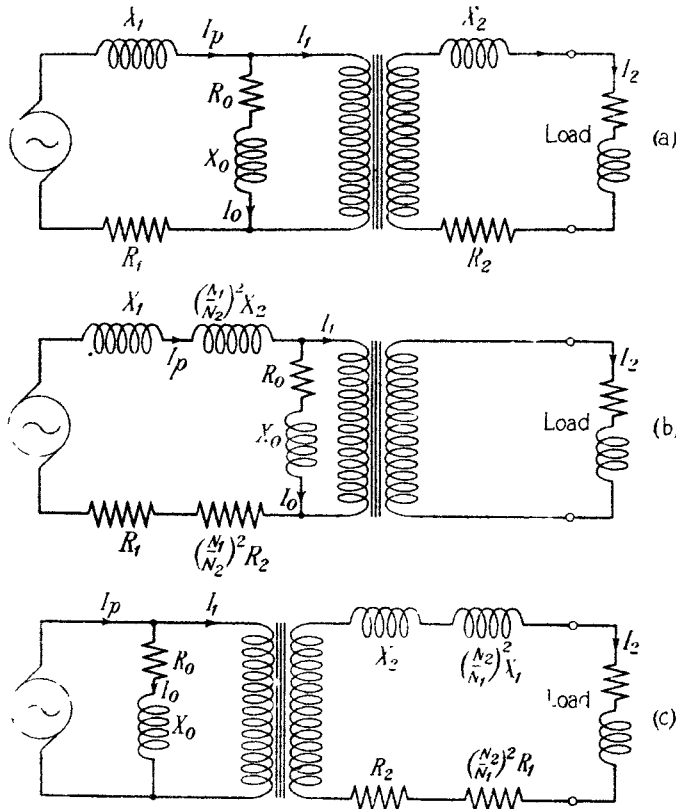


FIG. 14, CHAP. VI.—Equivalent circuits of transformer.

with the primary winding (fig. 14a). A further simplification may be made by using the equivalent primary resistance $R_p = R_1 + \frac{R_2}{T^2}$ and the equivalent leakage reactance $X_p = X_1 + \frac{X_2}{T^2}$ as in fig. 14b, or the equivalent secondary resistance $R_s = R_2 + T^2 R_1$ and equivalent secondary leakage reactance $X_s = X_2 + T^2 X_1$, fig. 14c. The value to be assigned to the quantities R_0 and X_0 can be determined by what is called the open circuit test upon the actual transformer. This is performed by connecting the transformer to supply mains of the voltage and frequency for which it was designed, a voltmeter (V_1), ammeter (A_1) and wattmeter (W) being included, as in fig. 15. The voltmeter (V_2) is inserted merely as a check on the transformation ratio and may be omitted if the turns ratio is known. In any event it should be of high resistance or of

CHAPTER VI.—PARA. 15

the electrostatic type. The wattmeter may be considered to give only the iron losses, because the copper losses are utterly insignificant under no-load conditions, while the ammeter (A_1) gives the no-load current, which may be divided into its two components, namely the iron-loss component I_i and true magnetising component I_m , by the following method. Referring to the vector diagram fig. 4 which shows the applied primary voltage V_p , the no-load current I_o and its components, it is seen that $I_i = I_o \cos \theta_o$ and $I_m = I_o \sin \theta_o$. Taking the 25 K.V.A. transformer previously referred to as an example, suppose the no-load current to be 2.5 amperes, and the wattmeter reading to be $P_i = 250$ watts. The power P_i is equal to $V_p I_o \cos \theta_o$, and therefore the iron loss component I_i of the no-load current is given by the equation $I_i = \frac{P_i}{V_p}$.

In this particular instance $I_i = \frac{250}{250} = 1$ ampere and the true magnetising component I_m is equal to $\sqrt{I_o^2 - I_i^2}$ or $\sqrt{2.5^2 - 1} = 2.265$ amperes. Hence $\tan \theta_o = \frac{I_m}{I_i} = 2.265$, and $\theta_o = 66^\circ$ approximately. It must not be supposed that these figures are representative of an efficient transformer.

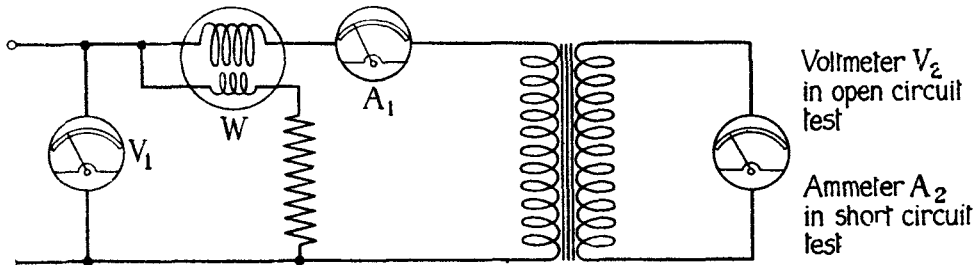


FIG. 15, CHAP. VI.—Arrangements for open circuit and short circuit tests of transformer.

The open circuit impedance Z_o of the primary is equal to $\frac{V_p}{I_o}$ or $\frac{250}{2.5} = 100$ ohms, and the resistance component of this is found by dividing the wattmeter reading by the square of the no-load current giving

$$R_o = \frac{250}{2.5^2} = \frac{250}{6.25} = 40 \text{ ohms.}$$

and the reactive component is now easily found because $Z_o = \sqrt{R_o^2 + X_o^2}$, whence

$$\begin{aligned} X_o &= \sqrt{Z_o^2 - R_o^2} \text{ or} \\ X_o &= \sqrt{100^2 - 40^2} \\ &= \underline{\underline{91.5 \text{ ohms.}}} \end{aligned}$$

15. The equivalent secondary resistance R_s and leakage reactance X_s are determined by means of the short circuit test. To perform this the secondary winding is short circuited by an ammeter (A_2), fig. 15, of very low resistance, and the primary winding is supplied at normal frequency, but with only sufficient terminal P.D. to cause full load secondary current to flow. The P.D. current and power supplied to the primary are then observed, by the instruments used in the previous test.

As the primary P.D. is low, only a comparatively weak flux is established and the iron losses are negligible compared with the copper losses. The wattmeter may therefore be considered to give the latter losses only.

Example.—The secondary of the transformer previously considered is connected to an ammeter of negligible resistance, and is found to give 10 amperes when the primary terminal

P.D. is 20 volts, the primary current 102.5 amperes and the input power 380 watts. Find the resistance and reactance of the transformer, assuming these are concentrated in the secondary winding.

$$\text{Equivalent secondary resistance } R_s = \frac{380}{10^2} = 3.8 \text{ ohms.}$$

The primary angle of lag is given by

$$\begin{aligned} \cos \theta_1 &= \frac{\text{Power}}{\text{volt-amperes}} \text{ in primary} \\ &= \frac{380}{102.5 \times 20} = .185 \\ \therefore \theta_1 &= 79^\circ \text{ (approx.).} \end{aligned}$$

The primary and secondary angles of lag are practically equal, because the secondary winding is giving its full load current.

$$\text{Hence } \tan \theta_2 = \tan \theta_1 = 5.3$$

$$\text{Now } \tan \theta_2 = \frac{X_s}{R_s}$$

$$\begin{aligned} \therefore X_s &= 5.3 R_s \\ &= 5.3 \times 3.8 \\ &= 20 \text{ ohms.} \end{aligned}$$

As the transformer has a 10 to 1 step-up, these values are 100 times as great as the equivalent primary resistance R_p and reactance X_p which become

$$R_p = .038 \text{ ohms.}$$

$$X_p = .2 \text{ ohms.}$$

Transformer efficiency

16. The efficiency of a transformer is measured by the ratio

$$\frac{\text{Power supplied to transformer—power wasted}}{\text{Power supplied to transformer}}$$

The power wasted is the sum of the copper and iron losses, the former increasing with the square of the secondary current and the latter being practically constant at all loads. If the copper losses at full load are known, e.g. as the result of a short circuit test, they are easily calculated for any other load, being $\frac{1}{4}$ of the maximum when the secondary current is half the full load, $\frac{1}{16}$ when the secondary current is one quarter full load and so on. In the 25 K.V.A. transformer which has been used in previous illustrations, the full load losses are, copper loss 380 watts, iron loss 250 watts, or a total of 630 watts, and the output is 25 K.V.A. If the secondary load is nonreactive, therefore, the input power for full load will be 25630 watts and the efficiency $\frac{25000}{25630} \times 100 = 97.5$ per cent.

At three-quarters full load the copper losses will be $(\frac{3}{4})^2 \times 380 = 213.75$ watts, the iron loss 250 watts, and the total losses $213.75 + 250 = 463.75$ watts. The output will be $\frac{3}{4} \times 25000 = 18750$ watts, and the efficiency $\frac{18750}{19194} = 97.6$ per cent.

At one-half full load the copper losses will be only 95 watts, and the total losses $95 + 250 = 345$ watts, while the output will be 12500 watts, hence

$$\eta = \frac{12500}{12845} = 97.2 \text{ per cent.}$$

CHAPTER VI.—PARA. 16

At one-quarter full load, again

copper losses = 24 watts (approx.)

iron losses = 250 watts

Total losses = 274 watts.

and $\eta = \frac{6250}{6524} = 95.7$ per cent.; at only 1/10 of full load the total losses are 254 watts, the output 2500 watts and $\eta = \frac{2500}{2754} = 90.7$ per cent.

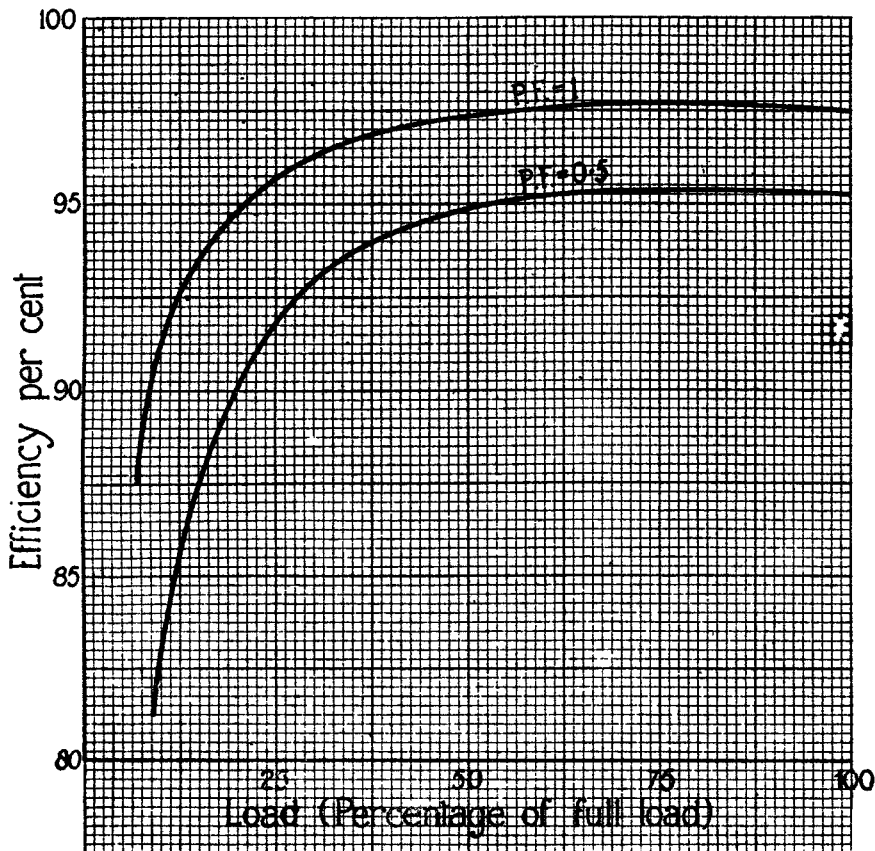


FIG. 16, CHAP. VI.—Variation of efficiency with load and power factor.

When the secondary load is reactive, the output in K.V.A. must be multiplied by the power factor of the load in order to obtain the output power; for example if the above calculations are repeated with a power factor of .5, the output watts are halved in each case, but the losses for each condition of loading are unchanged. At full load, therefore, the output is $25000 \times .5$ or 12500 watts, the input $12500 + 630$ watts and $\eta = \frac{12500}{13130} = 95.1$ per cent.

At half full load, the total losses are 345 watts, and the output 6250 watts.

$$\eta = \frac{6250}{6595} = 94.8 \text{ per cent.}$$

and so on. The efficiency at various degrees of loading and for the power factors 1.0 and 0.5 respectively are shown graphically in fig. 16.

It will be seen from the diagram that maximum efficiency does not occur at full load but at about 80 per cent. of full load. This is because maximum efficiency is achieved under loading conditions at which the copper and the iron losses are equal. If P_i is the iron loss and P_c the full load copper loss, the fraction of full load at which the efficiency is a maximum is $\sqrt{\frac{P_i}{P_c}}$ e.g. in the transformer dealt with in the above example, the full load copper loss is 380 watts, the iron loss 250 watts, and the efficiency is a maximum at $\sqrt{\frac{250}{380}} = .81$ of full load, the output being then 20,300 watts, the input 20,800 watts, and the efficiency approximately 97.6 per cent.

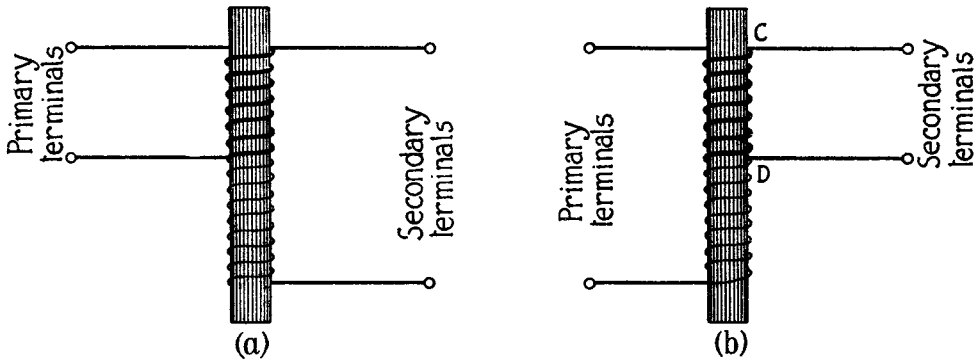


FIG. 17 CHAP. VI.—Auto-transformers.

The Auto-transformer

17. This is a transformer in which a portion of the winding is common to both primary and secondary circuits, and is only used where a low transformation ratio is required. Figs. 17a and 17b show the connections of the windings in the step-up and step-down types respectively, the portion of winding forming the primary being shown by a heavy line in the first instance, and the secondary winding by a heavy line in the second; thus the heavy line represents the conductor carrying the greatest current. The voltage applied to the primary winding causes a small magnetising current to flow, and a counter-E.M.F. of self-induction is set up in these turns which practically counterbalances the applied voltage. This current establishes in the core a magnetic flux which links with the whole of the secondary turns, and consequently induces in this winding an E.M.F. which is equal to $\frac{N_2}{N_1}$ or T times the applied voltage. When the secondary is on load, a secondary current will flow, tending to establish an additional flux in the core, which as in the ordinary transformer is neutralised by an equal and opposite flux set up by an increase of primary current, the latter being in antiphase with the secondary current which also flows in the common portion of the windings, and these turns therefore carry only the difference between the primary and secondary currents. This means that when only a small transformation ratio is desired the turns common to both windings need only be of a cross section sufficient to carry the difference between these two currents at full load. The copper losses are thus less than in a two-coil transformer designed for the same output and the same iron losses. The first cost of an

CHAPTER VI.—PARAS. 18-20

auto-transformer is also less than that of an ordinary transformer of the same output, owing to the saving of copper. These advantages are obviously only obtained when the primary and secondary currents are not very different, that is at low transformation ratios. The principal disadvantage of this type lies in the fact that there is direct connection between primary and secondary circuits. Thus in fig. 17b if a break occurs in the winding between C and D, the full primary voltage is applied to the device forming the secondary load, and the insulation of the latter will be damaged unless designed to stand this increased voltage.

Auto-transformers are used in the service to vary the voltage applied to the primary winding of the H.T. transformer of a rectifying system. Thus in the service Panel, Rectifying, Type A, which is designed for operation from a 230 volt supply, an auto-transformer having four secondary tapping points is provided, giving approximate step-down ratios of 2/1, 1.5/1, 1.2/1, and finally a 1/1 ratio, the corresponding voltages applied to the primary winding of the rectifier system being 110 volts, 150 volts, 190 volts and 230 volts.

Regulation

18. The regulation of a transformer is defined as the difference in terminal secondary voltage under full-load and no-load conditions, when the primary terminal P.D. is maintained at a constant value. That is

$$\text{Regulation} = E_s - V_s$$

The percentage regulation or "pressure rise" is the regulation expressed as a percentage of the secondary P.D. at full load or

$$\text{Percentage regulation} = \frac{E_s - V_s}{V_s} \times 100.$$

It should be noted that if the secondary load current is a leading one owing to the secondary external circuit possessing capacitive reactance, the secondary terminal P.D. on load may be higher than under no-load conditions.

Cooling

19. In order to radiate the heat generated in the core and windings either air or oil cooling is adopted. Air cooling is achieved by mounting the transformer in a protective casing of expanded metal, so that a free circulation of air may take place round the core and windings, and this circulation is assisted or increased if necessary by electric fans. Oil cooling consists of suspending the transformer in a tank containing insulating oil, which penetrates the windings and conducts the heat generated to the tank itself. The latter may be fitted with radiating fins in order that the heat may be rapidly radiated. In very large transformers the oil itself is cooled by water which circulates in pipes carried inside the tank. Such steps are not necessary in the comparatively small transformers used in the service, and plain oil-cooling is usually sufficient.

The purity of the oil used for cooling is of the utmost importance. It must contain neither dust particles, fluff, etc., nor water. Before filling a transformer tank it is advisable to stand the cans of oil in a warm place with the screw stopper removed for several hours and to strain the oil before filling the tank.

Extempore transformer design

20. (i) Although under ordinary conditions it is not necessary for transformer design to be undertaken by service personnel, circumstances may arise, particularly during hostilities, under which it is desirable to adapt an existing transformer for a voltage or frequency differing from its original rating, or even to construct a small transformer for temporary service pending the delivery of a correctly designed article. In the former event, a possible method of procedure is

outlined below, a concrete example being taken. It is required, then, to calculate the windings for a stalloy core having the dimensions given in fig. 18, to operate on 200 volt, 100 cycle mains, and to supply the following output :—

H.T. secondary, two windings each giving 100 milliamperes at 1000 volts.

L.T. secondary, two windings each giving 10 amperes at 15 volts.

The total output is therefore

H.T. :— $1000 \times .1 \times 2 = 200$ volt-amperes.

L.T. :— $15 \times 10 \times 2 = 300$ volt-amperes.

Total 500 volt-amperes.

For simplicity the power factor of the load will be assumed to be unity and the power output 500 watts.

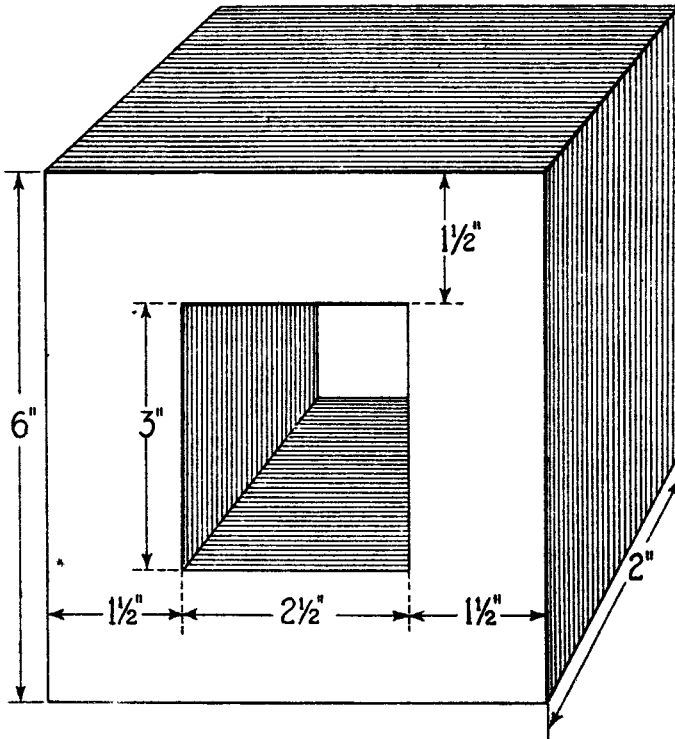


FIG. 18, CHAP. VI.—Dimensions of core.

(ii) The first step is to decide upon a tentative value for the peak flux density $\mathcal{B}$. This may be found as follows. Assume an efficiency which is reasonably capable of fulfilment, say 94 per cent. and calculate the total losses on full load. If P_0 is the output, P_1 the input, $P_1 - P_0$ the losses and η the efficiency, $P_1 = \frac{P_0}{\eta}$ in the particular example $P_1 = 500 \div 0.94 = 532$ watts, and the losses will be 32 watts ; if the copper and iron losses are equal the iron losses will be 16 watts. Weigh the core, excluding clamping plates and bolts, or calculate its weight by finding the total volume of the iron in cubic inches and multiplying by the weight of one cubic inch of iron, 0.28 lb. The cross section of the core is $1\frac{1}{2}$ in. $\times$ 2 in. but of this a proportion is

CHAPTER VI.—PARA. 20

insulating material between stampings; allowing ten per cent. for this the iron cross-section (A) is $0.9 \times 3 = 2.7$ sq. in. and the volume 51 cubic in. of which 0.9 is iron. The approximate amount of the iron is therefore $46 \times .28 = 13$ lb. and the iron loss will be $\frac{16}{13} = 1.23$ watts per lb. If the thickness of the stampings is .03 inch reference to fig. 22 at the end of this section shews that the peak flux density should not exceed 6,700 gauss.

(iii) Next, the voltage induced in each turn of the windings must be found. Since $e = \frac{Nd\phi}{10^8 dt}$, and the flux is sinusoidal, $\frac{e}{N} = \frac{\omega\phi_{\max}}{10^8}$, e being the peak voltage. What is actually required is the R.M.S. voltage per turn or $\frac{E}{N}$, and $E = \frac{e}{\sqrt{2}}$ hence

$$\begin{aligned}\frac{E}{N} &= \frac{.707 \omega A \mathcal{B}}{10^8} \\ &= \frac{4.44 f A \mathcal{B}}{10^8}\end{aligned}$$

This is the fundamental formula in transformer design. In this formula A must be expressed in square centimetres. In the present example, $A = 2.7$ sq. in. or 2.7×6.45 sq. cm. and

$$\begin{aligned}\frac{E}{N} &= 4.44 \times 6700 \times 2.7 \times 6.45 \times 100 \times 10^{-8} \\ &= .520 \text{ volts per turn.}\end{aligned}$$

The number of turns required for each winding now follows :—

Primary voltage	200	
Primary turns	$\frac{200}{.52}$	= 390.
H.T. secondary voltage	= 1000	
H.T. secondary turns	= 390×5	= 1950
L.T. secondary voltage	= 15	
L.T. secondary turns	= $\frac{15}{200} \times 390$	= 30

$\left. \begin{array}{l} \text{H.T. secondary voltage} = 1000 \\ \text{H.T. secondary turns} = 390 \times 5 = 1950 \\ \text{L.T. secondary voltage} = 15 \\ \text{L.T. secondary turns} = \frac{15}{200} \times 390 = 30 \end{array} \right\} \begin{array}{l} \text{per winding.} \\ \text{per winding.} \end{array}$

N.B.—The number of turns on each winding should be a whole number.

(iv) The wire gauges may now be decided upon. Commencing with the primary, the full-load current will be $\frac{532 \text{ watts}}{200 \text{ volts}} = 2.66$ amperes, and the current density in the copper should

not exceed 1,000 amperes per square inch. Referring to Table I Appendix A it is seen that at this density, No. 17 s.w.g. is hardly capable of carrying 2.66 amperes while No. 16 s.w.g. will be operated at less than the above density and will therefore develop less heat. The even number wire gauges are in more general use than the odd numbers and 16 s.w.g. is more likely to be available than No. 17. If the number of volts per turn is less than four, single cotton covered wire may be used, although double cotton covered is preferable for the primary and silk-covered wire for the H.T. secondary. The space required for the actual winding is found from Table XI, Appendix A. It is there stated that No. 16, s.c.c., gives 198 turns per square inch of winding space. The 390 primary turns therefore require 1.97 sq. in. Each L.T. secondary is to carry 10 amperes and from the abovementioned tables it is found that No. 10 s.w.g. is suitable and single cotton covered wire of this gauge takes 54 turns per square inch. Each winding then

occupies $\frac{30}{54} = .556$ sq. in. It will be convenient to divide the primary winding into two equal portions, one on each limb, and to arrange one-half of the primary, one L.T. secondary and one H.T. secondary on each limb of the core. The space occupied so far is 1.97 sq. in. for the primary and 1.112 sq. in. for the two L.T. secondaries, a total of 3.082 sq. in. As the total window space is $2\frac{1}{2}$ in. $\times$ 3 in. = 7.5 sq. in. this leaves say 4.4 sq. in. for the H.T. secondary and for insulation.

The H.T. secondary must carry 0.1 ampere and a suitable wire is No. 30 s.w.g. (d.s.c.). This gives 4,500 turns per sq. in. and each winding will occupy $\frac{1950}{4500} = .432$ sq. in. or a total of .864 sq. in. There is therefore ample space for insulation between windings. It is of interest to recalculate the space occupied by the windings, assuming that only d.c.c. wire is available. It is then found that the space available for insulation is rather small but that the number of turns of given gauges could be accommodated, provided great care is taken to form the coils correctly and to use the minimum of tape binding on them. The suggested arrangement is shewn diagrammatically in figs. 19 and 20. In fig. 19, the additional insulation at the angles of the core should

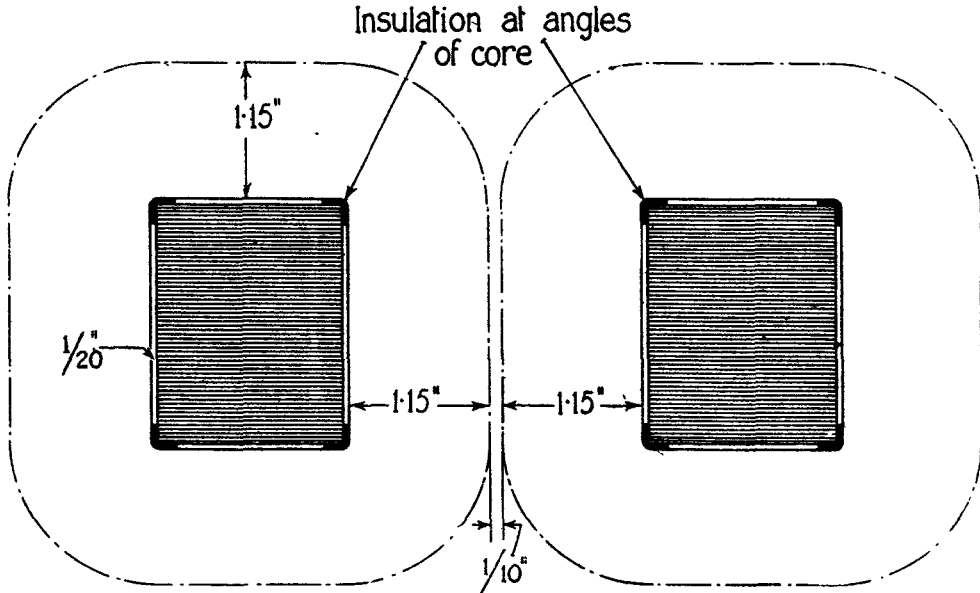


FIG. 19, CHAP. VI.—Sectional plan of core and windings.

be noted ; it is at these points that breakdown of the insulation between the H.T. secondary and the core is most liable to occur. In fig. 20, it will be observed that the windings have been sectionalised to a greater extent than suggested above, each H.T. secondary being divided into six sections so that the peak P.D. between the ends of each section is less than 240 volts. Each half of the primary is also divided into two sections and intermingled with the secondaries in order to reduce magnetic leakage as much as possible.

21. (i) Before proceeding further it is advisable to calculate the temperature rise which will take place if 32 watts are dissipated. For this purpose the following empirical formula may be used :—

$$\theta = \frac{250 (P_i - P_o)}{\text{Total cooling surface in sq. in.}}$$

where θ is the temperature rise in degrees centigrade. In calculating the cooling surface, the superficial area of the windings and the iron are added, but parallel surfaces within 0.5 inch of each other should be omitted as experience shews that such surfaces contribute little or nothing to the cooling. On this basis, the cooling area, calculated from figs. 19 and 20, will be about 160 sq. in. and the temperature rise

$$\theta = \frac{250 \times 32}{160} = 50^\circ \text{ C.}$$

CHAPTER VI.—PARA. 21

Cotton covering begins to char at 85° C. and in commercial design the final temperature is not allowed to exceed 80° C. If the normal temperature is 15° C., the final temperature reached by the transformer under discussion will be 65° C. which is perfectly satisfactory.

(ii) Now calculate the copper losses at 65° C. The temperature coefficient of copper is .004 and the resistance of all windings will be 20 per cent. greater than at 15° C. The length of the "mean turn" is about one foot, and the following results are easily obtained.

Primary 390 turns = 130 yds. = .97 ohm at 15° C. = 1.17 ohms at 65° C.

H.T. secondary 3900 „ = 1300 „ = 254 „ „ „ = 305 „ „ „

L.T. „ 60 „ = 20 „ = .037 „ „ „ = .045 „ „ „

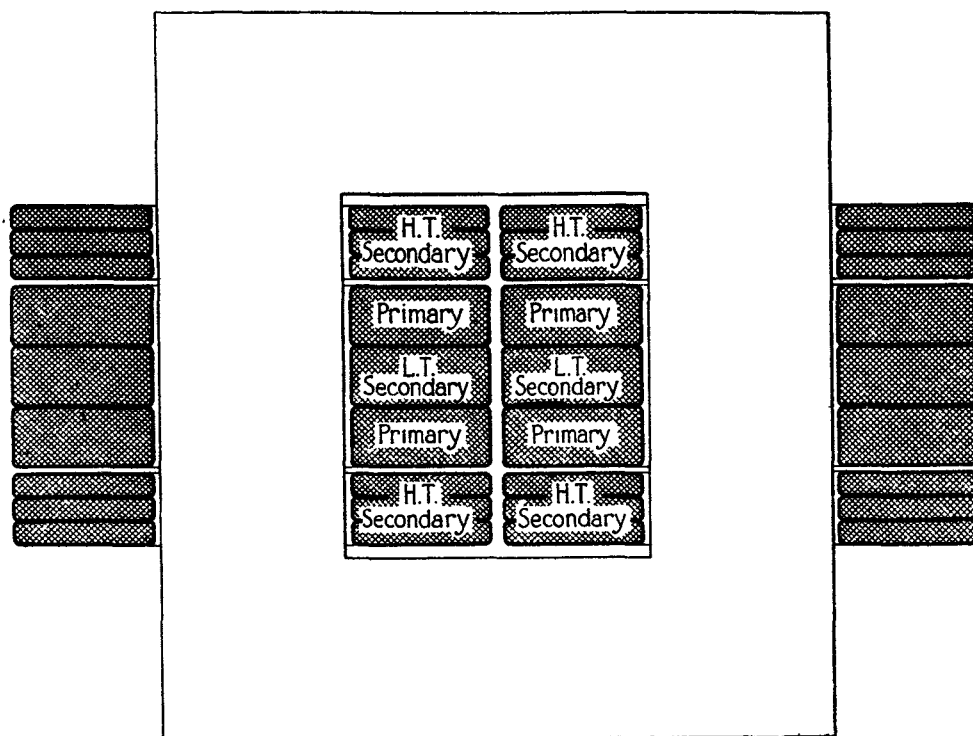


FIG. 20, CHAP. VI.—Sectional elevation of core and windings.

Copper losses.

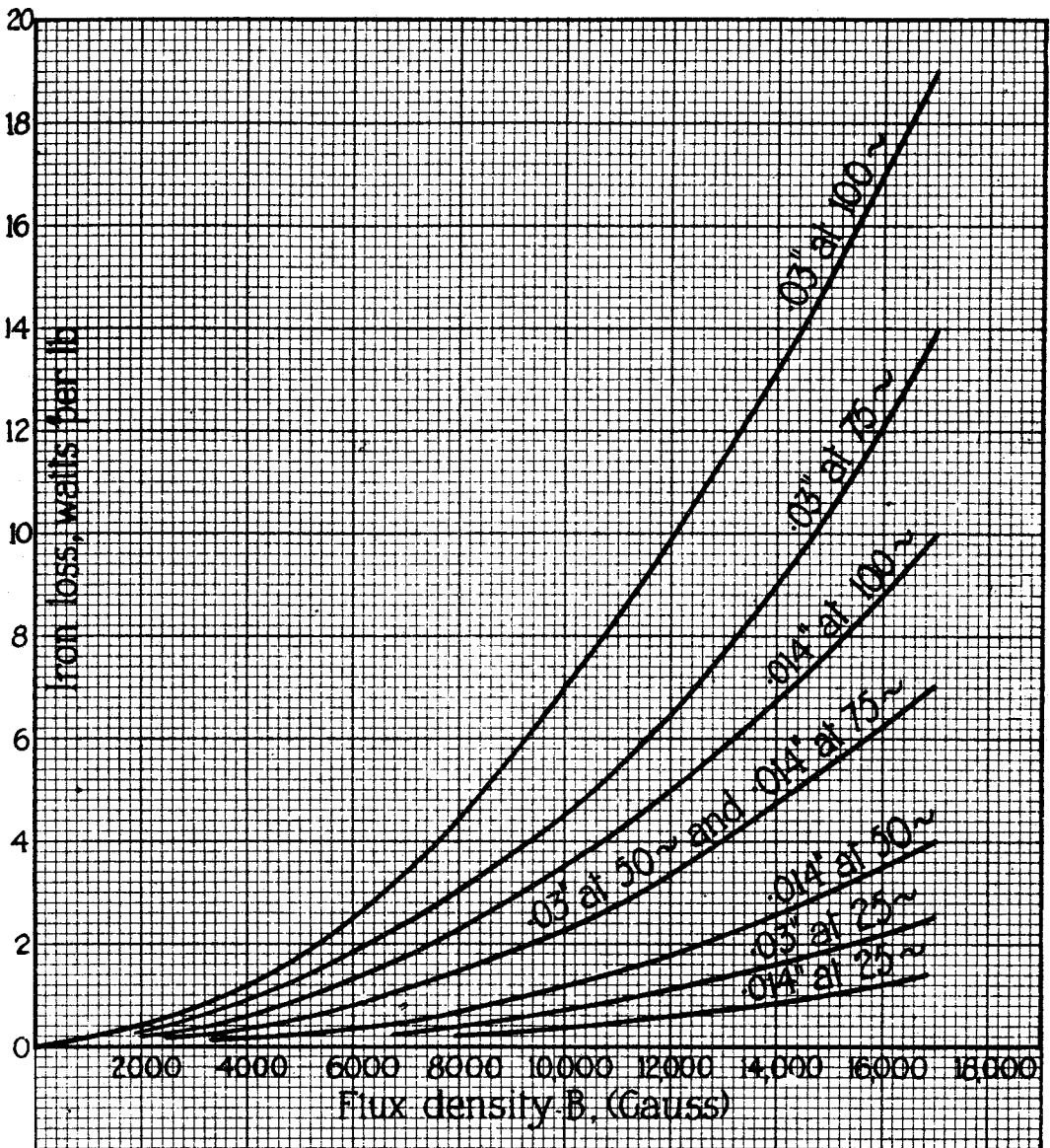
Primary $2.7^2 \times 1.17 = 8.5$ watts.

H.T. secondary $.1^2 \times 305 = 3.05$ „

L.T. „ $10^2 \times .045 = 4.5$ „

16.05 watts

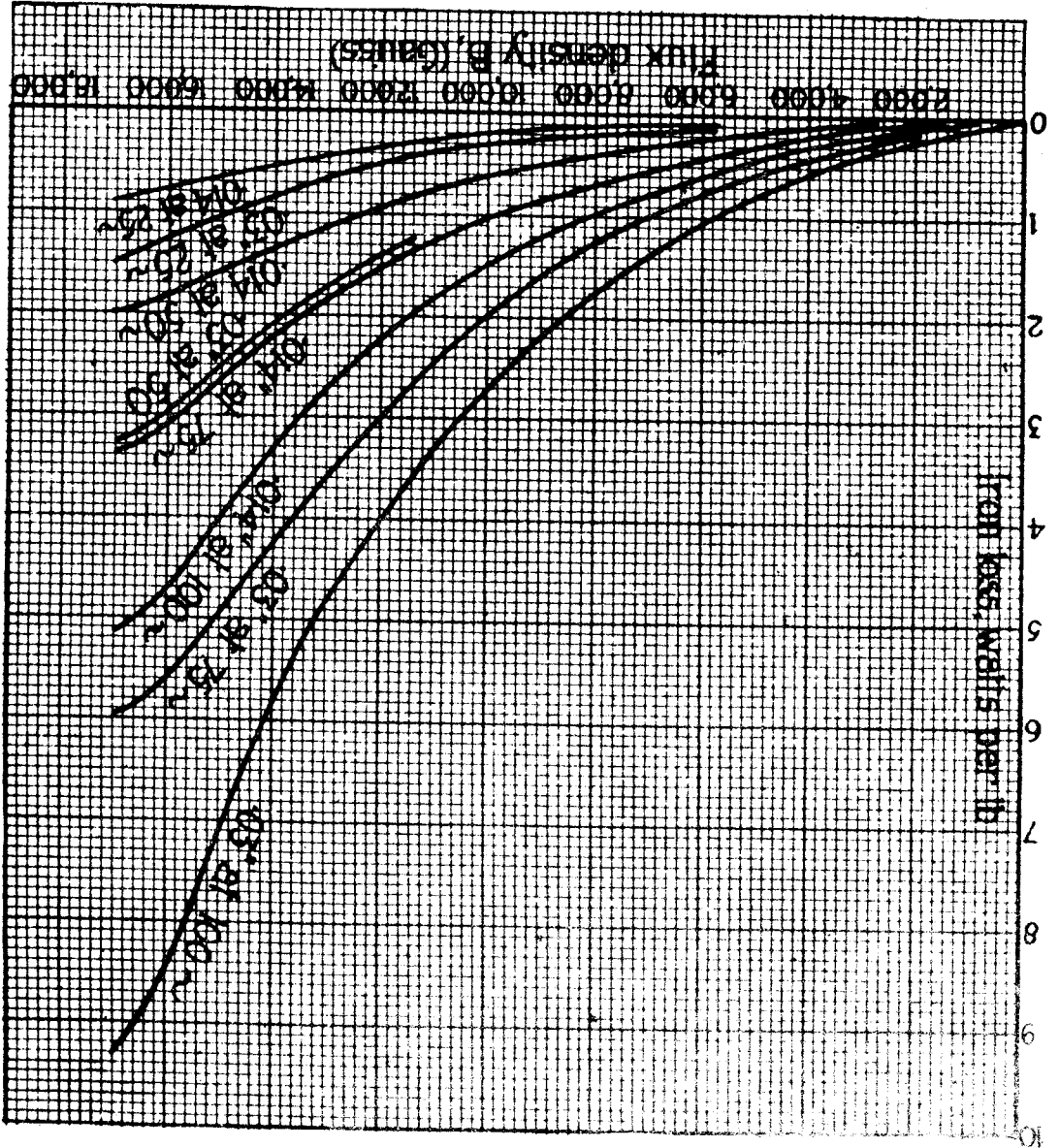
The primary current is taken as 2.7 amperes instead of 2.66 in order to allow for the magnetising current.

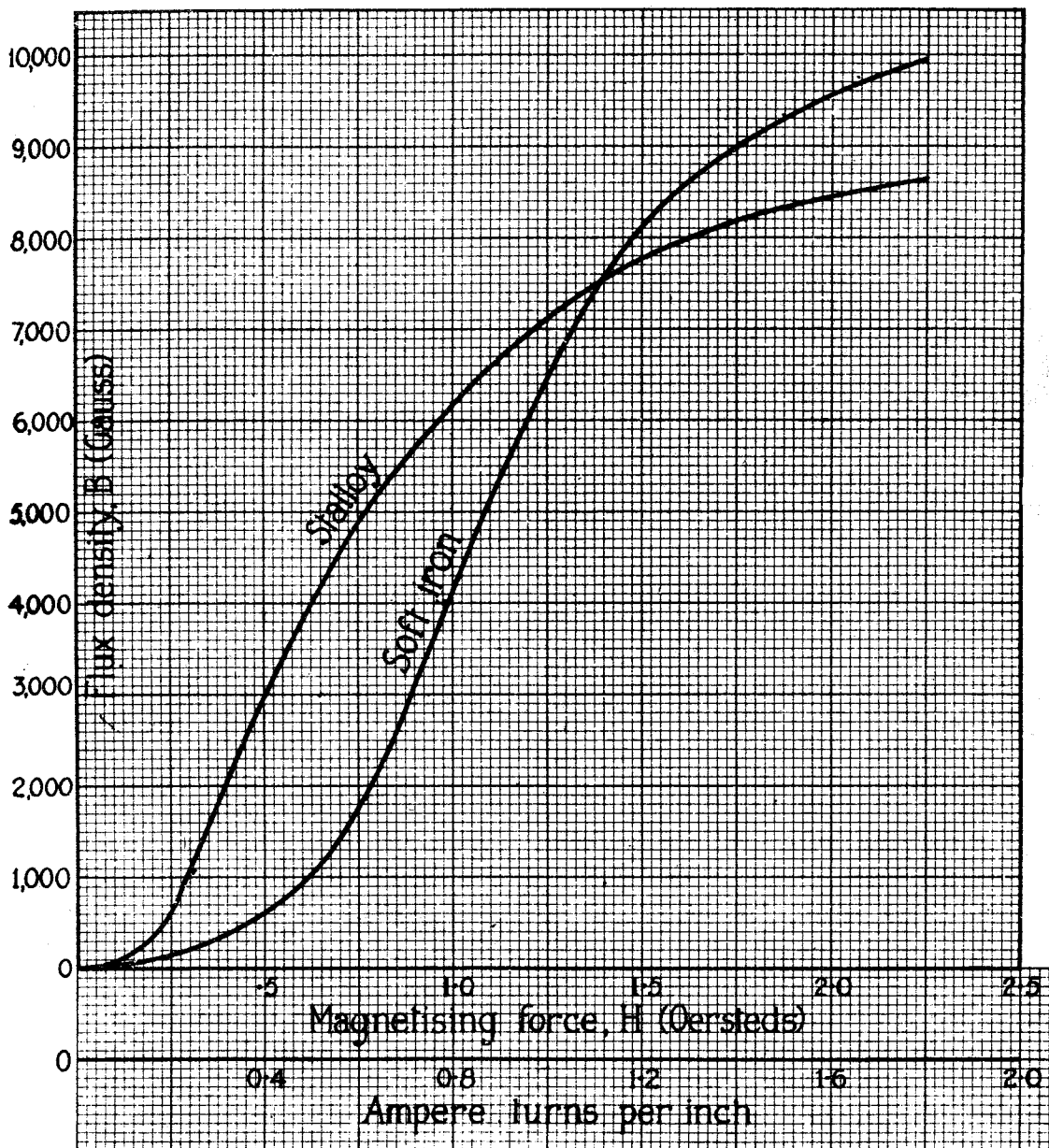


**VARIAION OF IRON LOSS WITH FLUX DENSITY
(SOFT IRON STAMPINGS)**

**FIG.2I
CHAP.VI**

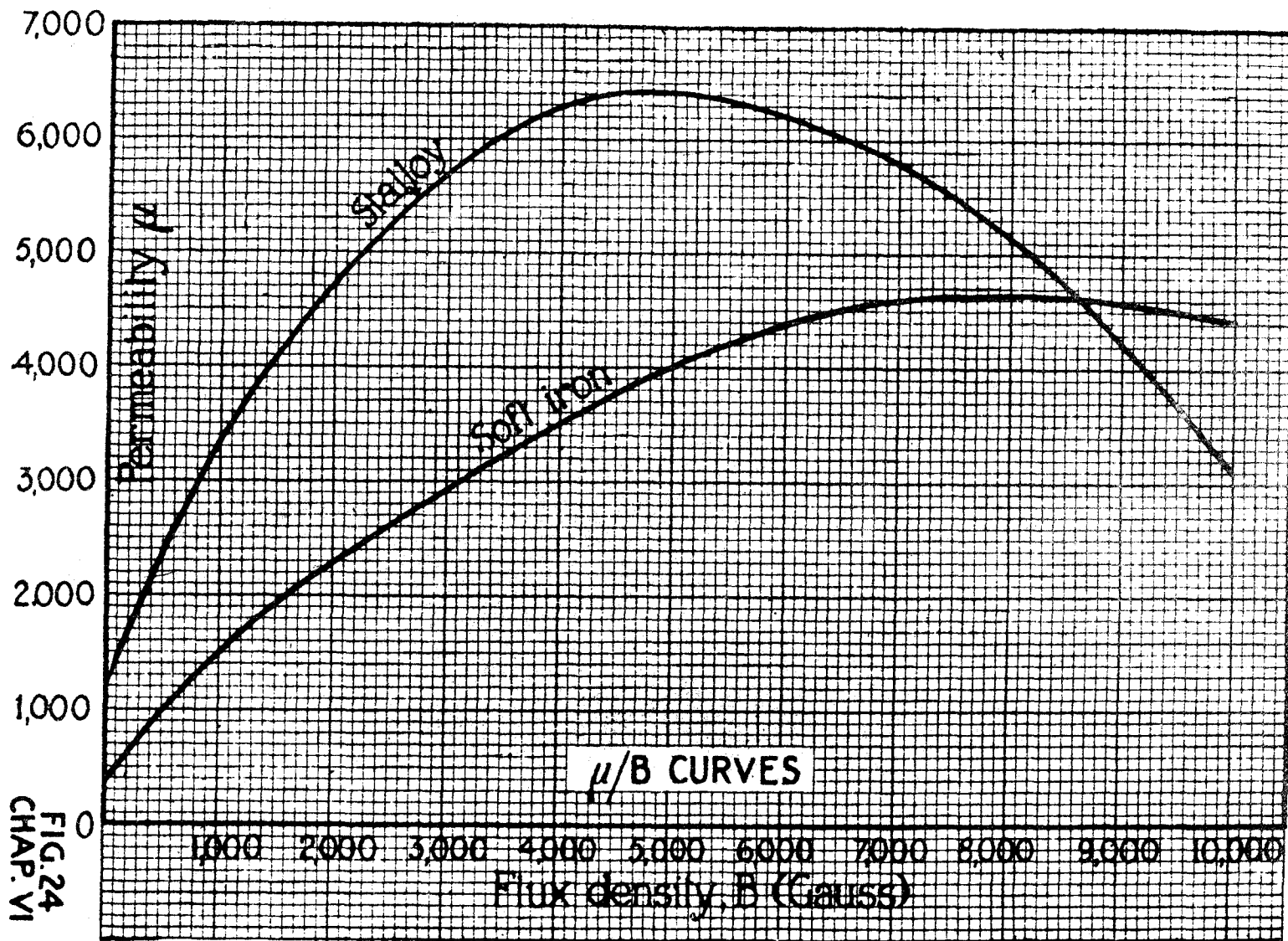
VARIATION OF IRON LOSS WITH FLUX DENSITY (STALLOY STAMPINGS)





B/H CURVES

**FIG.23
CHAP.VI.**



(iii) The regulation cannot be calculated accurately because the primary and secondary leakage inductances are not known, but assuming that the leakage is negligible the volts drop in the primary winding will be $2.7 \times 1.2 = 3.24$ volts, in the H.T. secondary $.1 \times 305 = 30.5$ volts and in the L.T. secondary $10 \times .045 = .45$ volts. The terminal P.D.'s on full load are then easily found.

$$\text{Primary terminal P.D.} = 200 - 3.24 = 196.76 \text{ volts.}$$

$$\begin{aligned} \text{H.T. secondary E.M.F.} &= 196.76 \times 10 \\ &= 1967.6 \text{ volts.} \end{aligned}$$

$$\begin{aligned} \text{H.T. secondary, terminal P.D.} &= 1967.6 - 30.5 \\ &= 1937.1 \text{ volts.} \end{aligned}$$

$$\text{L.T. secondary E.M.F.} = \frac{196.76 \times 15}{200} = 14.76 \text{ volts.}$$

$$\text{L.T. secondary, terminal P.D.} = 14.76 - .45 = 14.31 \text{ volts.}$$

22. When a small transformer has to be improvised, the only core material usually available is ordinary soft iron sheet or tinplate. Provided it is carefully annealed from a suitably high temperature, and operated at low flux density, such material is fairly satisfactory. The simplest procedure is to choose dimensions which appear to be reasonable and then calculate the performance. An assumption of core cross-section and flux density will determine the volts per turn and hence the number of turns on each winding, the size of wire and winding space. The core itself may be built up from long strips which are threaded through the windings and bent round to form a closed magnetic circuit.

Example.—A transformer is required to supply 10 amperes at 20 volts from 200 volt 50 cycle mains. Soft iron of .03 in. thickness is available.

$$\text{Assume an efficiency of 91 per cent. The input is then } \frac{200}{.91} = 222 \text{ watts.}$$

$$\text{Losses} = 22 \text{ watts.}$$

$$\text{Assume, e.g., iron loss} = 12 \text{ watts.}$$

$$\text{copper loss} = 10 \text{ watts.}$$

From fig. 21 a suitable flux density will be 6,000 gauss, giving an iron loss of .8 watt per lb., and $\frac{12}{.8} = 15$ lb. of iron is required, the volume being $15 \div .28 = 53.5$ or say 54 cu. in.

A gross core cross section of 2 in. $\times$ 2 in. or 4 sq. in. and a mean flux line of 15 in. giving a gross volume of 60 cu. in., will give a window of $1\frac{1}{2}$ in. $\times$ 2 in. The iron cross-section will be only $.9 \times 4 = 3.6$ sq. in. and the iron volume 54 cu. in., as already stated.

The volts per turn will be

$$\frac{E}{N} = \frac{4 \times 1.11 \times 50 \times 6000 \times 3.6 \times 6.45}{10^8}$$

$$= \frac{2 \times 1.11 \times 6 \times 3.6 \times 6.45}{1000}$$

$$= .31 \text{ volts/turn}$$

CHAPTER VI.—PARAS. 23-24

$$\text{Primary turns } \frac{200}{.31} = 645.$$

$$\text{Secondary turns } \frac{20}{200} \times 645 = 64.5.$$

Primary current 1 ampere 20 s.w.g., 567 turns per sq. in.

Secondary current 10 amperes 10 s.w.g., 54 turns per sq. in.

$$\left. \begin{array}{l} \text{Primary winding space} \quad \frac{645}{567} = 1.14 \text{ sq. in.} \\ \text{Secondary winding space} \quad \frac{64.5}{54} = 1.2 \text{ sq. in.} \end{array} \right\} \text{Total, } 2.34 \text{ sq. in.}$$

Single silk covered wire being used in both windings.

Thus ample room for insulation is available in the window space as the highest voltage is less than 300. The mean turn will be say 11 inches in length. Roughing out the dimensions it is estimated that the cooling surface will be about 125 sq. in. and the temperature rise

$$\frac{250}{125} \times 22 = 44^\circ \text{C.}$$

Taking the mean turn as one foot the total length of primary conductor is 645 feet or 215 yds. and its resistance $.215 \times 23.54 \times 1.8 = 9.2$ ohms. The length of the secondary winding will be about 21.5 yds., and its resistance $.0215 \times 1.862 \times 1.8 = .0723$ ohms.

The copper losses in the primary will be $1^2 \times 9.2 = 9.2$ watts and in the secondary, $10^2 \times .0723 = 7.23$ watts, hence the total copper loss will be 13.43 watts.

This is sufficiently near the required performance to justify the adoption of the suggested core dimensions, and the arrangements of core and windings may now be drawn with accuracy, the length of mean turn, the cooling surface, and the copper losses recomputed. Finally the magnetising current may be estimated as in the previous instance.

COUPLED CIRCUITS

23. When two circuits are so arranged that electrical energy can be transferred from one circuit to another owing to the existence of some form of impedance which is common to both circuits, the latter are said to be coupled together and are briefly designated as coupled circuits. Coupled circuits are usually divided into three classes, according to the nature of the common impedance, giving rise to (i) magnetic or inductive coupling (ii) electric or capacitive coupling (iii) resistive coupling. The most familiar case of magnetic coupling is that of the power transformer already considered, but coupled circuits are also of considerable importance in radio-frequency practice, both in transmission and reception. The individual circuits are referred to as the primary and secondary respectively, and usually possess both inductance and capacitance in addition to inherent resistance; the primary and secondary circuits are usually tuned to the same frequency.

Mutual inductive coupling

24. (i) This is shewn diagrammatically in fig. 25a. The primary circuit consists of the inductance L_p , the resistance R_p , and the capacitance C_p , and energy is supplied by an alternator having an R.M.S. voltage of E volts. The secondary circuit consists of the inductance L_s , capacitance C_s and resistance R_s , and the common impedance is that due to the mutual flux-linkage between L_p and L_s , its magnitude at any frequency $\frac{\omega}{2\pi}$ being ωM . When an E.M.F. is applied to the primary circuit an alternating current I_1 will be established and will set up an alternating flux round the coil I_1 . This varying flux will embrace the inductance L_s setting up and alternating E.M.F. in the coil and consequently a current in the secondary circuit. By Lenz's law, this secondary flux must act upon the primary circuit in such a manner as to oppose its original

cause, hence the secondary current induces in the primary circuit a counter-E.M.F. The presence of the secondary circuit is therefore responsible for the following phenomena. First, the current in the primary circuit is not the same as it would be in the absence of the secondary current, for since the primary E.M.F. causes a current in the secondary circuit, additional energy must be supplied by the alternator, and consequently the effective impedance of the two coupled circuits must be different from that of the primary alone. Second, the resonant frequency of the whole circuit is modified by the presence of the secondary. The manner in which the latter effect arises may be shewn as follows. Let the primary E.M.F. be E , and the frequency of the alternator $\frac{\omega}{2\pi}$

Then assuming an alternating current I_1 to be established in the primary circuit, the secondary induced voltage will be $E_2 = \omega M I_1$. If the reactances of the primary and secondary circuits are X_p and X_s respectively, and the resistances R_p , R_s are negligible, the secondary current will

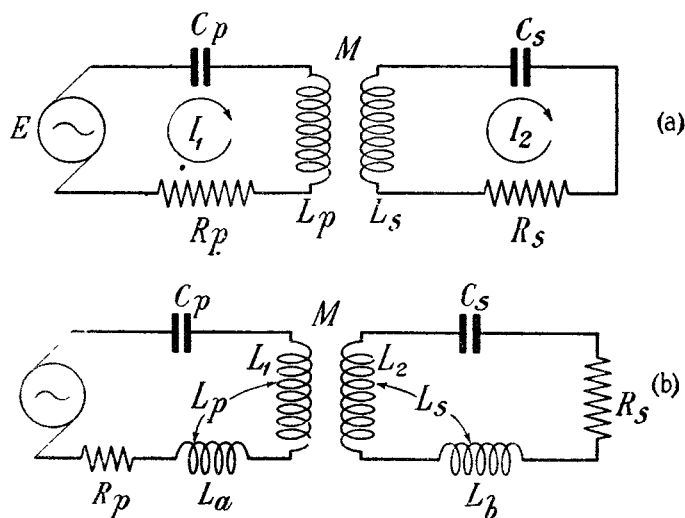


FIG. 25, CHAP. VI.—Mutual inductive coupling.

be $\frac{E_2}{X_s}$ and $I_2 = \frac{\omega M}{X_s} I_1$; this secondary current in its turn induces an E.M.F. in the inductance L_p , the value of which is $\omega M I_2$ or $\frac{\omega^2 M^2 I_1}{X_s}$ so that the total E.M.F. which is acting in the primary circuit is not E but $E + \frac{\omega^2 M^2}{X_s} I_1$.

(ii) The magnitude of the primary current can now be found, for it is equal to the total E.M.F. divided by the opposition of the circuit or

$$I_1 = \frac{E + \frac{\omega^2 M^2}{X_s} I_1}{X_p}$$

which simplifies to $X_p I_1 = E + \frac{\omega^2 M^2}{X_s} I_1$

$$I_1 \left(X_p - \frac{\omega^2 M^2}{X_s} \right) = E$$

or finally

$$I_1 = \frac{E}{X_p - \frac{\omega^2 M^2}{X_s}}$$

CHAPTER VI.—PARA. 25

The denominator $X_p - \frac{\omega^2 M^2}{X_s}$ is the ratio of voltage to current in the primary circuit,

that is, its apparent reactance, and this has been modified from its original value X_p by the presence of the secondary circuit. The resonant frequency of the whole circuit can now be established, for it is that frequency which makes the reactance zero, that is

$$X_p - \frac{\omega^2 M^2}{X_s} = 0$$

i.e. $X_p X_s - \omega^2 M^2 = 0.$

Now for simplicity assume that the two circuits are identical so that $L_p = L_s$, $C_p = C_s$, $X_p = X_s$,

$$= \omega L_p - \frac{1}{\omega C_p}, X_p X_s = \left(\omega L_p - \frac{1}{\omega C_p} \right)^2, \text{ and}$$

$$\left(\omega L_p - \frac{1}{\omega C_p} \right)^2 - \omega^2 M^2 = 0$$

$$\omega^2 L_p^2 - \frac{2L_p}{C_p} + \frac{1}{\omega^2 C_p^2} - \omega^2 M^2 = 0$$

or $(L_p^2 C_p^2 - M^2 C_p^2) \omega^4 - 2 L_p C_p \omega^2 + 1 = 0$

which may be treated as a quadratic equation. Solving it to find ω^2 ,

$$\begin{aligned} \omega^2 &= \frac{2 L_p C_p \pm \sqrt{(2 L_p C_p)^2 - 4 (L_p^2 - M^2) C_p^2}}{2 (L_p^2 - M^2) C_p^2} \\ &= \frac{2 L_p C_p \pm 2 M C_p}{2 (L_p^2 - M^2) C_p^2} \\ &= \frac{L_p \pm M}{(L_p^2 - M^2) C_p}. \end{aligned}$$

Now $L_p^2 - M^2 = (L_p + M) (L_p - M)$ and therefore the two values of ω^2 given by this equation are

$$\omega_1^2 = \frac{1}{(L_p + M) C_p}$$

$$\omega_2^2 = \frac{1}{(L_p - M) C_p}.$$

25. (i) In order that the values of ω_1 and ω_2 may be conveniently expressed without introducing the absolute value of the mutual inductance M , it is now desirable to introduce the "coefficient of coupling," or "coupling factor." The latter is defined algebraically as the ratio

$$\frac{M}{\sqrt{L_p L_s}} \text{ and is denoted by the symbol } k, \text{ hence in the particular instance where } L_p = L_s, k = \frac{M}{L_p},$$

and

$$\omega_1^2 = \frac{1}{L_p C_p (1 + k)}$$

$$\omega_2^2 = \frac{1}{L_p C_p (1 - k)}$$

It will be found that the same result is obtained even if L_p is not equal to L_s , provided that $L_p C_p = L_s C_s$. If the latter relation is not satisfied, the circuits as a whole still have two resonant frequencies, but the condition is of little practical importance. The resonant frequencies of the circuits are directly derived from the values of ω_1^2 and ω_2^2 , and are

$$f_1 = \frac{\omega_1}{2\pi} = \frac{1}{2\pi\sqrt{L_p C_p (1 + k)}} = \frac{f_r}{\sqrt{1 + k}}$$

$$f_2 = \frac{\omega_2}{2\pi} = \frac{1}{2\pi\sqrt{L_p C_p (1 - k)}} = \frac{f_r}{\sqrt{1 - k}}$$

where f_r is the resonant frequency of each individual circuit, $\frac{1}{2\pi\sqrt{L_p C_p}}$ or $\frac{1}{2\pi\sqrt{L_s C_s}}$.

(ii) The value of the coefficient of coupling depends upon the ratio of the flux linking with both circuits to that linking with the individual circuits. It has been shown in Chapter II that if the whole of the flux embraces every portion of both circuits $M = KN_p N_s$, $L_p = KN_p^2$, $L_s = KN_s^2$, N_p and N_s being the number of turns in the primary and secondary inductances and K a constant depending upon the geometry of the circuits. In these circumstances

$$k = \frac{M}{\sqrt{L_p L_s}} = \frac{KN_p N_s}{\sqrt{KN_p^2 \times KN_s^2}} = 1$$

i.e. the coefficient of coupling is unity. As it is impossible to achieve the totality of mutual flux-linkage specified above, the constant k can never reach the value unity in any practical circuit, although it is approached closely if the inductances L_p and L_s are intermingled on a common iron core as in the power transformer. The coefficient of coupling can be reduced by any method which reduces the amount of mutual flux linkage, e.g. by separating the circuits in space, using a core of non-magnetic material, and by turning the coils into such a relative position that the flux due to the primary current does not link with any portion of the secondary circuit. When calculating the coefficient of coupling between two circuits, some care is necessary, for instance in circuits arranged as in fig. 25b. The total primary inductance in this arrangement is $L_p = L_a + L_1$ and the secondary inductance $L_s = L_b + L_2$. Let us suppose that the mutual inductance between L_1 and L_2 has been measured, and its value known to be M . Then the

coupling factor for the coils L_1, L_2 alone is $k = \frac{M}{\sqrt{L_1 L_2}}$ but for the circuits as a whole is $\frac{M}{\sqrt{L_p L_s}}$,

which may be very much less. An example of this is a type of air-core transformer in which the secondary winding L_s possesses a large number of turns, while the primary winding L_p consists of a few turns wound closely over one end of the secondary coil. The mutual inductance between L_p and the few secondary turns immediately underneath may be very nearly equal to L_p , i.e. the coupling factor between L_p and an equal inductance forming part of L_s may be unity. The

coupling factor for the whole circuit will then be $\frac{M}{\sqrt{L_p L_s}} = \frac{L_p}{\sqrt{L_p L_s}} = \sqrt{\frac{L_p}{L_s}}$, and when two coils

are arranged in this way, the coupling factor decreases with an increase of secondary inductance, i.e. with the number of turns wound on the secondary, and consequently an increase of turns ratio in order to achieve a larger step-up of voltage may be stultified by the reduction of the coupling factor.

CHAPTER VI.—PARA. 26

Example.—In fig. 25b, the coefficient of coupling between the coils L_1 and L_2 is $\cdot 8$. If $L_1 = 50\mu H$, $L_a = 100\mu H$, $L_2 = 70\mu H$, $L_b = 200\mu H$, find the coefficient of coupling between the two circuits.

$$\begin{aligned}\frac{M}{\sqrt{L_1 L_2}} &= \cdot 8 \\ M &= \cdot 8 \sqrt{L_1 L_2} \\ &= \cdot 8 \sqrt{50 \times 70} \mu H \\ k &= \frac{M}{\sqrt{L_p L_s}} \\ &= \frac{\cdot 8 \sqrt{50 \times 70}}{\sqrt{(100 + 50)(200 + 70)}} \\ &= \frac{\cdot 8 \sqrt{3500}}{\sqrt{40500}} \\ &= \cdot 236.\end{aligned}$$

Resonance curves of coupled circuits

26. (i) When two circuits (individually tuned to the same frequency) are coupled together by mutual induction and an E.M.F. of variable frequency is applied, the graphical representation of the variation of current with frequency, over a band extending below and above the resonant frequency, is called the resonance curve of the coupled circuits. It is perhaps obvious that separate curves may be drawn shewing the variation of current in the primary and secondary circuits respectively, but the secondary current is of principal interest in practice, and only this current will be dealt with. When the coupling factor is very low, e.g. of the order of $\cdot 002$, the peak value secondary current at the resonant frequency is very small, and the ratio $\frac{I_r}{I_n}$ is large

(I_r being the current at resonant and I_n the current at any other frequency) but as the coefficient of coupling is increased, a condition is reached in which the mutual reactance ωM is equal to the geometric mean resistance of the circuits, $\sqrt{R_p R_s}$. This is termed the critical coupling, because for this value of the mutual inductance M the secondary current reaches maximum value, being then given by the equation

$$I_2 (\max.) = \frac{\omega M E}{\omega^2 M^2 + R_p R_s}$$

Since

$$\omega M = \sqrt{R_p R_s}$$

$$I_2 (\max.) = \frac{E}{2\sqrt{R_p R_s}}$$

This is the greatest current which can be obtained in the secondary circuit. An increase in the value of M does not give a further increase in secondary current, but results in the formation of two peaks in the resonance curve, the frequencies at which these peaks appear being given by

the formulæ $f_1 = \frac{f_r}{\sqrt{1+k}}$, $f_2 = \frac{f_r}{\sqrt{1-k}}$ previously obtained. The frequencies f_1 and f_2 become

more widely separated as k is increased, fig. 26 shewing the resonance curves of two circuits, the constants being the same as in the single circuit used to illustrate the simple resonance curve of the acceptor circuit (Chapter V) namely $L_p = L_s = 150 \mu H$, $C_p = C_s = .000169 \mu F$, $R_p = R_s = 9.45$ ohms. approx., $f_r = 1000$ k.c/s. A separate curve has been shewn for each of several values of k , and it will be observed that an increase of the coupling coefficient, from $k = .002$ upwards results in an increase of secondary current until k reaches a value .01. Further increase in the coupling simply has the effect of broadening the resonance curve without increasing the value of the current, the curve becoming double peaked as above stated. The maximum transfer of energy at the resonant frequency f_r will occur when the coefficient of coupling has the critical

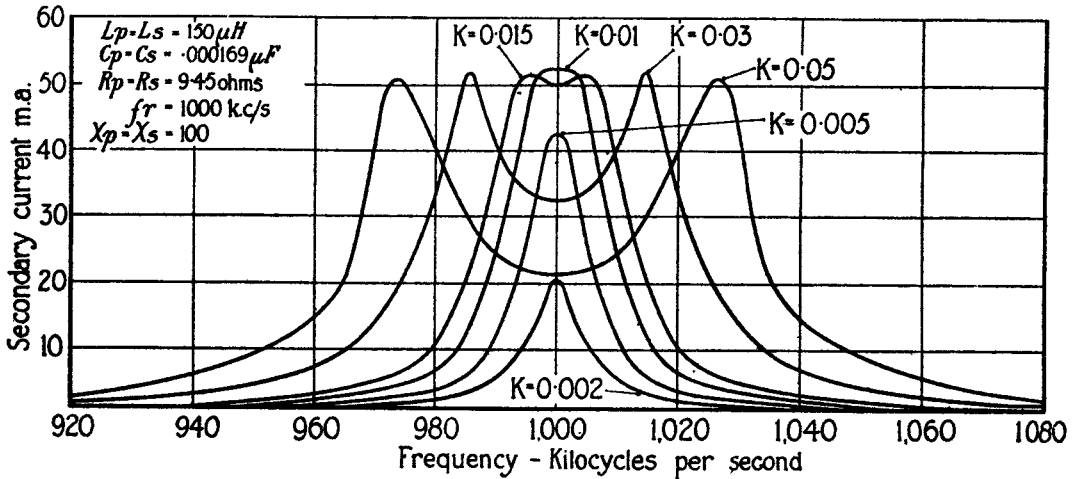


FIG. 26, CHAP. VI.—Resonance curves of coupled circuits.

value given by the relation $\omega^2 M^2 = R_p R_s$. As the value of M is not usually known it is preferable to derive an expression for the critical value of the coupling factor k , as follows :—

$$\omega^2 M^2 = R_p R_s$$

$$M^2 = \frac{R_p}{\omega} \frac{R_s}{\omega}$$

$$\frac{M^2}{L_p L_s} = \frac{R_p}{\omega L_p} \frac{R_s}{\omega L_s}$$

Since

$$\frac{M^2}{L_p L_s} = k^2, \frac{R_p}{\omega L_p} = \frac{1}{x_p}, \frac{R_s}{\omega L_s} = \frac{1}{x_s},$$

$$k = \sqrt{\frac{1}{x_p} \times \frac{1}{x_s}}$$

This is the value of coupling coefficient for which the transfer of energy at the resonant frequency is a maximum, and as in practical radio frequency circuits x_p and x_s may be of the order of 100 or more, it is apparent that very loose coupling is generally sufficient to secure this optimum transference.

CHAPTER VI.—PARA. 27

(ii) Referring again to the equations connecting the coupling factor k , the resonant frequency f_r and the frequencies f_1 and f_2 at which peaks occur in the resonance curve, it will be observed that if the curve is obtained experimentally, the coefficient of coupling can be calculated from the frequencies f_1 and f_2 . Thus, since

$$f_1 = \frac{f_r}{\sqrt{1+k}}, f_2 = \frac{f_r}{\sqrt{1-k}}, f_1^2 + f_1^2 k = f_r^2, f_2^2 - f_2^2 k = f_r^2,$$

and $f_1^2 + f_1^2 k = f_2^2 - f_2^2 k$.

Hence $f_2^2 - f_1^2 = (f_1^2 + f_2^2)k$

and $k = \frac{f_2^2 - f_1^2}{f_2^2 + f_1^2}.$

In fig. 26, the frequencies at which peaks occur, on the curve showing the highest degree of coupling, are approximately $f_1 = 975$ kilocycles per second, $f_2 = 1,025$ kilocycles per second. Hence

$$k = \frac{1025^2 - 975^2}{1025^2 + 975^2} = \frac{105 \cdot 0625 - 94 \cdot 9625}{105 \cdot 0625 + 94 \cdot 9625} = \frac{10 \cdot 1}{200 \cdot 025}$$

$$\therefore k = \cdot 05005$$

It will also be observed that the frequencies f_1 and f_2 differ from the resonant frequency of the individual circuits by a nearly equal amount, namely 25 kilocycles per second below and above the resonant frequency; this signifies that when the coupling is of the order usually employed, that is if k is less than about $\cdot 1$, the above formula may be replaced by a simpler one with negligible error, namely

$$k = \frac{f_2 - f_1}{f_r}$$

Using this approximation in the above example

$$k = \frac{1025 - 975}{1000} \\ = \cdot 05$$

Auto-inductive coupling

27. This is shewn in fig. 27, the notation being similar to that used in previous figures.

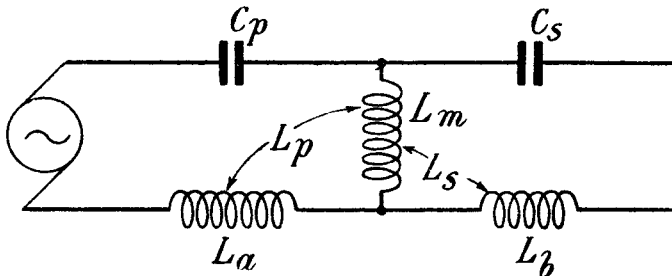


FIG. 27, CHAP. VI.—Auto-inductive coupling.

The reactance which is common to both primary and secondary circuits is that of the inductance L_m . The primary inductance is $L_p = L_a + L_m$ and the secondary inductance $L_s = L_b + L_m$.

If the circuits are so adjusted that $L_p C_p = L_s C_s$, the circuit as a whole has two resonant frequencies, which are given by the expressions

$$f_1 = \frac{1}{2\pi\sqrt{L_p C_p (1+k)}}$$

$$f_2 = \frac{1}{2\pi\sqrt{L_p C_p (1-k)}}$$

and

$$k = \frac{L_m}{\sqrt{L_p L_s}}$$

The degree of coupling may therefore be varied by variation in the value of the inductance L_m , but, except in the special case when $L_a = L_b$, $C_p = C_s$, such a variation will throw the circuits out of resonance with each other. If, however, the circuits are so designed that $L_b = FL_a$ and $C_a = FC_b$ (where F is any numeric whatever) the two condensers being mounted in such a way that they are varied simultaneously by a single knob, then for all settings of the condenser the following relation is satisfied, viz. $L_a C_p = L_b C_s$. The circuit as a whole still possesses two resonant frequencies, but instead of being above and below the resonant frequency of the individual circuits, they are

$$f_1 = \frac{1}{2\pi\sqrt{L_a C_p \left(1 + \frac{L_m}{L_a} + \frac{L_m}{L_b}\right)}}$$

$$f_2 = \frac{1}{2\pi\sqrt{L_a C_p}}$$

If the value of the inductance L_m is varied while the rest of the circuit constants remain at given values, the resonant frequency f_2 remains constant but the frequency f_1 decreases with increase of L_m , and the resonance curve is said to have a stationary peak at f_2 and a moving peak at the frequency f_1 which depends upon the degree of coupling. If the coefficient of coupling is defined by the equation

$$k = \frac{f_2^2 - f_1^2}{f_2^2 + f_1^2} = \frac{\omega_2^2 - \omega_1^2}{\omega_2^2 + \omega_1^2},$$

and the above expressions for ω_1^2 and ω_2^2 are inserted,

$$\begin{aligned} k &= \frac{\frac{1}{L_a C_p} - \frac{1}{L_a C_p \left(1 + \frac{L_m}{L_a} + \frac{L_m}{L_b}\right)}}{\frac{1}{L_a C_p} + \frac{1}{L_a C_p \left(1 + \frac{L_m}{L_a} + \frac{L_m}{L_b}\right)}} \\ &= \frac{L_m \left(\frac{1}{L_a} + \frac{1}{L_b}\right)}{2 + L_m \left(\frac{1}{L_a} + \frac{1}{L_b}\right)} \\ &= \frac{L_m}{2 \frac{L_a L_b}{L_a + L_b} + L_m} \end{aligned}$$

CHAPTER VI.—PARA. 27

Example.—(i) In a certain circuit $L_a = 140 \mu H$, $C_p = .001 \mu F$, $L_m = 20 \mu H$, $L_b = 300 \mu H$, $C_s = .0005 \mu F$. Find the two resonant frequencies, and the coefficient of coupling.

Since $L_p = L_a + L_m = 140 + 20 = 160 \mu H$, $L_s = L_b + L_m = 300 + 20 = 320 \mu H$, $L_p C = 160 \times .001$, $L_s C_s = 320 \times .0005$, $L_p C_p = L_s C_s$ and the coefficient of coupling is

$$k = \frac{L_m}{\sqrt{L_p L_s}} = \frac{20}{\sqrt{160 \times 320}} \\ = .0883.$$

The resonant frequency of each individual circuit is

$$f_r = \frac{1}{2\pi \sqrt{L_p C_p}} = \frac{10^6}{2\pi \sqrt{160 \times .001}} \\ = \frac{10^6}{2\pi \times .4} = 398000 \text{ cycles per second.} \\ f_1 = \frac{f_r}{\sqrt{1+k}} = \frac{398000}{\sqrt{1.0883}} = 381000 \text{ cycles per second.} \\ f_2 = \frac{f_r}{\sqrt{1-k}} = \frac{398000}{\sqrt{.9117}} = 417000 \text{ cycles per second.}$$

(ii) If the secondary inductance is reduced to $280 \mu H$ the remainder of the constants being unchanged, find the resonant frequencies and coefficient of coupling.

$L_a C_p$ is now equal to $L_b C_s$ and the system has a stationary and a variable resonant frequency.

$$f_1 = \frac{10^6}{2\pi \sqrt{140 \times .001 \left(1 + \frac{20}{140} + \frac{20}{280}\right)}} \\ = \frac{10^6}{2\pi \sqrt{.14 \left(1 + \frac{3}{14}\right)}} \\ = 385000 \text{ cycles per second.} \\ f_2 = \frac{10^6}{2\pi \sqrt{.14}} \\ = 424000 \text{ cycles per second.} \\ k = \frac{L_m}{2 \frac{L_a L_b}{L_a + L_b} + L_m} \\ = \frac{20}{2 \frac{140 \times 280}{420} + 20} \\ = .0967.$$

Alternatively, the approximation $k = 2 \frac{f_2 - f_1}{f_2 + f_1}$ gives

$$k = 2 \frac{424 - 385}{809} \\ = \cdot 0964.$$

Auto-capacitive coupling

28. This is shewn in fig. 28 and is analogous to that just discussed. The primary capacitance

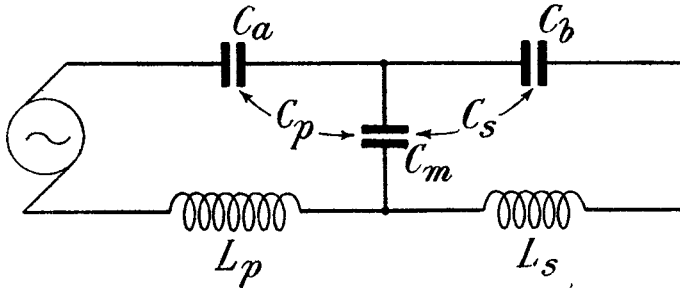


FIG. 28, CHAP. VI.—Auto-capacitive coupling.

is $C_p = \frac{C_a C_m}{C_a + C_m}$ and the secondary capacitance $C_s = \frac{C_b C_m}{C_b + C_m}$. If the circuits are so arranged that $L_p C_p = L_s C_s$, the circuit possesses two resonant frequencies, viz. :

$$f_1 = \frac{\sqrt{1 - k}}{2\pi \sqrt{L_p C_p}}$$

$$f_2 = \frac{\sqrt{1 + k}}{2\pi \sqrt{L_p C_p}}$$

where

$$k = \sqrt{\frac{C_a C_b}{(C_a + C_m)(C_b + C_m)}}$$

This condition is not normally maintained because unless $C_a = C_b$, an alteration in the coupling by adjustment of the capacitance C_m will throw the circuits out of resonance with each other. The more usual arrangement is to make $L_p C_a = L_s C_b$. The system then has one stationary resonant frequency.

$$f_1 = \frac{1}{2\pi \sqrt{L_p C_a}}$$

and a resonant frequency varying with the degree of coupling.

$$f_2 = \frac{1}{2\pi \sqrt{L_p C_a \left(1 - \frac{C_a + C_b}{C_a + C_b + C_m}\right)}}$$

CHAPTER VI.—PARA. 29

Under these circumstances, if the coefficient of coupling is defined by the equation

$$k = \frac{f_2^2 - f_1^2}{f_2^2 + f_1^2}$$

$$k = \frac{C_a + C_b}{C_a + C_b + 2C_m}$$

Example.—(i) If $L_p = 160 \mu H$, $C_a = .00111 \mu F$, $L_s = 320 \mu H$, $C_b = .000526 \mu F$, $C_m = .01 \mu F$, find the resonant frequencies and coefficient of coupling.

$$C_p = .001 \mu F, C_s = .0005 \mu F, L_p C_p = L_s C_s$$

$$k = \sqrt{\frac{1.111 \times 10^{-3} \times 5.26 \times 10^{-4}}{1.111 \times 10^{-3} \times 1.0526 \times 10^{-2}}}$$

$$= \sqrt{\frac{5.84 \times 10^{-7}}{1.168 \times 10^{-4}}}$$

$$= .0707.$$

$$f_1 = \frac{.9293 \times 10^6}{2\pi\sqrt{.16}} = 369000 \text{ cycles per second.}$$

$$f_2 = \frac{1.0707 \times 10^6}{2\pi\sqrt{.16}} = 426000 \text{ cycles per second.}$$

(ii) If C_b is increased to .000555, so that $L_p C_a = L_s C_b$, find the resonant frequencies and coefficient of coupling.

$$f_1 = \frac{1}{2\pi\sqrt{L_p C_a}}$$

$$= \frac{10^6}{2\pi\sqrt{160 \times .00111}} = 378000 \text{ cycles per second.}$$

$$f_2 = \frac{1}{2\pi\sqrt{L_p C_a \left(1 - \frac{C_a + C_b}{C_a + C_b + C_m}\right)}}$$

$$C_a + C_b = .001666$$

$$C_a + C_b + C_m = .011666$$

$$1 - \frac{C_a + C_b}{C_a + C_b + C_m} = 1 - .1428 = .8572$$

$$f_2 = \frac{10^6}{2\pi\sqrt{160 \times .00111 \times .8572}}$$

$$= 410000 \text{ cycles per second.}$$

Direct inductive coupling

29. This is shewn in fig. 29, but has only been mentioned for sake of completeness, as it is rarely used in practice, for two reasons. First, in order to attain the low degree of coupling generally required, the inductance L_m must be very much larger than that of the coils L_p , L_s , and must also be adjustable to within fairly fine limits, requirements which are incompatible

with each other. Second, such a large inductance must of necessity possess considerable self-capacitance, and the circuit as a whole does not function in the manner calculated on the assumption that the self-capacitance is zero.

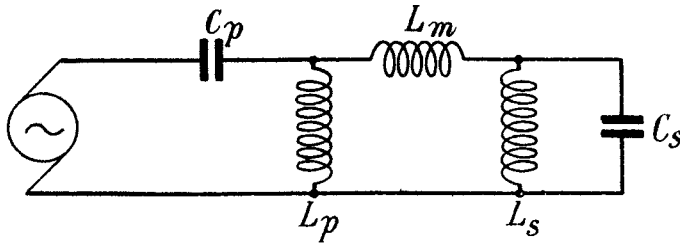


FIG. 29, CHAP. VI.—Direct inductive coupling.

Direct capacitive coupling

30. This type of circuit is analogous to that mentioned above, but a small condenser C_m constitutes the coupling device. This is one of the most convenient types of coupling and is probably in more general use than any other with the exception of mutual inductive coupling. The circuits are invariably so adjusted that $L_p C_p = L_s C_s$ and the circuit possesses a stationary frequency

$$f_2 = \frac{1}{2\pi\sqrt{L_p C_p}}$$

and a variable frequency

$$f_1 = \frac{1}{2\pi\sqrt{L_p C_p \left(1 + \frac{C_m}{C_p} + \frac{C_m}{C_s}\right)}}$$

which depends upon the degree of coupling.

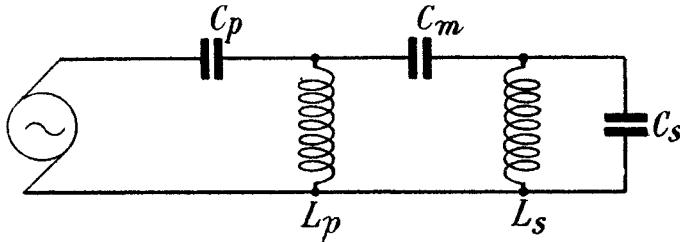


FIG. 30, CHAP. VI.—Direct capacitive coupling.

The coupling factor is found as before :—

$$\begin{aligned} k &= \frac{f_1^2 - f_2^2}{f_1^2 + f_2^2} \\ &= \frac{C_m}{2 \frac{C_p C_s}{C_p + C_s} + C_m} \end{aligned}$$

If $C_p = C_s$ this reduces to $\frac{C_m}{C_p + C_m}$. In practice the capacitance of the coupling condenser is always very small, e.g. if $C_p = C_s = .0005 \mu F$, and it is desired to maintain a coefficient of coupling of .01 between the two circuits, the appropriate value of C_m is only .000005 μF (approximately).

CHAPTER VI.—PARA. 31

Resistive coupling

31. In fig. 31 are shewn two oscillatory circuits having a resistance R_m common to both, and by analogy with the circuits previously discussed, this arrangement may be referred to as auto-resistive coupling. It is rarely adopted deliberately for the purpose of energy transference,

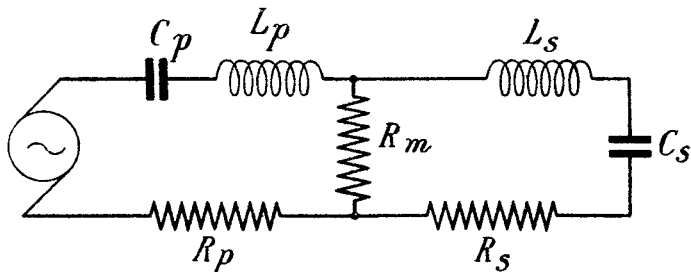


FIG. 31, CHAP. VI.—Auto-resistive coupling.

for the introduction of additional resistance must of necessity increase the damping and reduce the selectivity of the individual circuits, but such coupling may be found to exist fortuitously, for instance where earth connections are made to different points between which the resistance is appreciable. If both circuits are tuned to the same frequency and an E.M.F. of this frequency is applied, the reactance of each circuit is zero and the primary and secondary currents may be found in exactly the same manner as in D.C. practice, giving

$$I_1 = \frac{R_s + R_m}{R_p R_s + R_m(R_s + R_p)} E$$

$$I_2 = \frac{R_m}{R_p R_s + R_m(R_s + R_p)} E$$

It is of interest to calculate the relative magnitude of the primary and secondary currents in circumstances which might be found in practice. Suppose $R_p = R_s = 10$ ohms, $R_m = 0.1$ ohm, and the applied E.M.F. to be 1 volt. Application of the above formulae then gives

$$I_1 = \frac{10.1}{102} \div 0.1 \text{ ampere, } I_2 = \frac{0.1}{102} \div 0.001 \text{ ampere.}$$

Hence the presence of the resistance R_m gives rise to a secondary current equal in amplitude to one per cent. of that in the primary circuit, and the circuits are sometimes said to have a coupling factor of 0.01. It must be appreciated, however, that as the presence of the coupling resistance has no effect upon the resonant frequency, the resonance curve of the combination possesses only a single peak, and in the true sense of the term the circuits do not possess a coupling factor. Fig. 32 shews a second form of resistive coupling which may be found to exist; it may be referred to as direct resistive coupling. In this instance the apparent coupling factor increases with decrease of the value of the coupling resistance R_m . The arrangement is rarely adopted for the reasons stated with regard to the alternative form of resistive coupling.

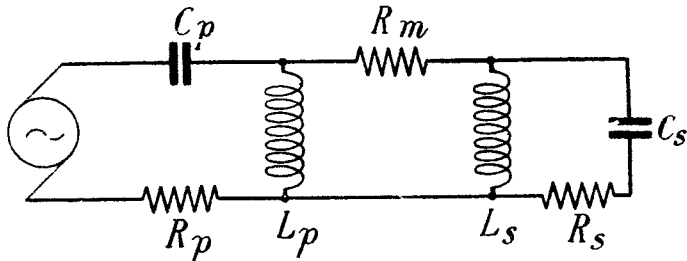


FIG. 32, CHAP. VI.—Direct resistive coupling.

CHAPTER VII.—ELECTRICAL OSCILLATIONS AND ELECTROMAGNETIC WAVES

MECHANICAL AND ELECTRICAL OSCILLATIONS**Mechanical equivalents of inductance, resistance and capacitance**

1. In the first two chapters the charge and discharge of a condenser through a resistance, and the growth and decay of current through a circuit containing inductance and resistance, were individually discussed. In radio circuits, the properties of resistance, inductance and capacitance are generally found in combination, and a knowledge of the phenomena associated with the discharge of a condenser in a circuit possessing both inductance and resistance is a fundamental requirement in the study of the principles of radio communication. The first portion of this chapter is therefore devoted to a consideration of the properties of oscillatory circuits, i.e. those containing inductance, capacitance and resistance, in which no continuously applied electromotive force exists.

2. An excellent grasp of the principles involved in the oscillatory circuit can be obtained by the study of a mechanical analogy, which unlike most analogies is practically perfect, namely, the mechanical vibration of a body or system of connected bodies, possessing the properties of mass, friction and elasticity. It has already been stated that the effect of inductance in an electrical circuit is to oppose any change in the value of the current flowing; in this respect inductance resembles that property of a body which we call its mass, for the distinctive property of mass is inertia, or opposition to any change in the motion of a body. If the body is at rest, it can only be set in motion by the application of a force, and if the force acts only for a short interval of time the body tends to continue in motion with the velocity it possessed at the instant at which the force is removed. An example of this may be seen when railway trucks laden with coal are being moved from a siding in the vicinity of the pithead to the railway proper, for which purpose horses are usually employed. Two or more horses may be required to urge one truck into motion, but when sufficient velocity has been attained the horses are dispensed with and the truck maintained in motion with nearly constant velocity by the unaided effort of one man. It will also be observed that to bring the truck to a sudden standstill considerable force must be applied. These results are very noticeable when trucks run on smooth well-laid lines, because the friction between wheels and rails is very small, and the effort required to maintain constant velocity is much less than is required upon a rough surface such as a highway.

3. The effect of friction in mechanics is generally similar to the effect of resistance in electrical circuits, but friction between solid surfaces is not an exact parallel to electrical resistance, the latter being more nearly analogous to the friction which exists between a smooth body and a viscous fluid when a body moves slowly through the latter, for in these circumstances the velocity produced is proportional to the force applied, or, if F is the force and u the velocity produced, $F \propto u$ or $F = Ru$. The constant of proportion R , which has been introduced to give equality to both members of the equation, may be called the coefficient of friction. This equation may be compared with Ohm's law which is $E = RI$, where E is the applied E.M.F., R the electrical resistance, and I the current. It will be noted that the current I is correctly compared with the velocity u , because the intensity of the current is the rate at which electrons move through the circuit, being measured in coulombs per second (one coulomb = 6.29×10^{18} electrons). It is because of this analogy that practical men often think of the E.M.F. as the force acting upon the electrons, although we have seen in Chapter I that a more accurate conception of E.M.F. is based upon the conversion of energy into its electrical form.

4. Reverting to the mechanical analogue of inductance, namely, inertia, it may be recalled that when a body is in motion it possesses kinetic energy, the quantity of energy being proportional to the mass of the body and the square of its velocity. Care must be taken not to confuse the mass of the body, which is merely the amount of matter it contains, with its weight, which is

CHAPTER VII.—PARAS. 5-6

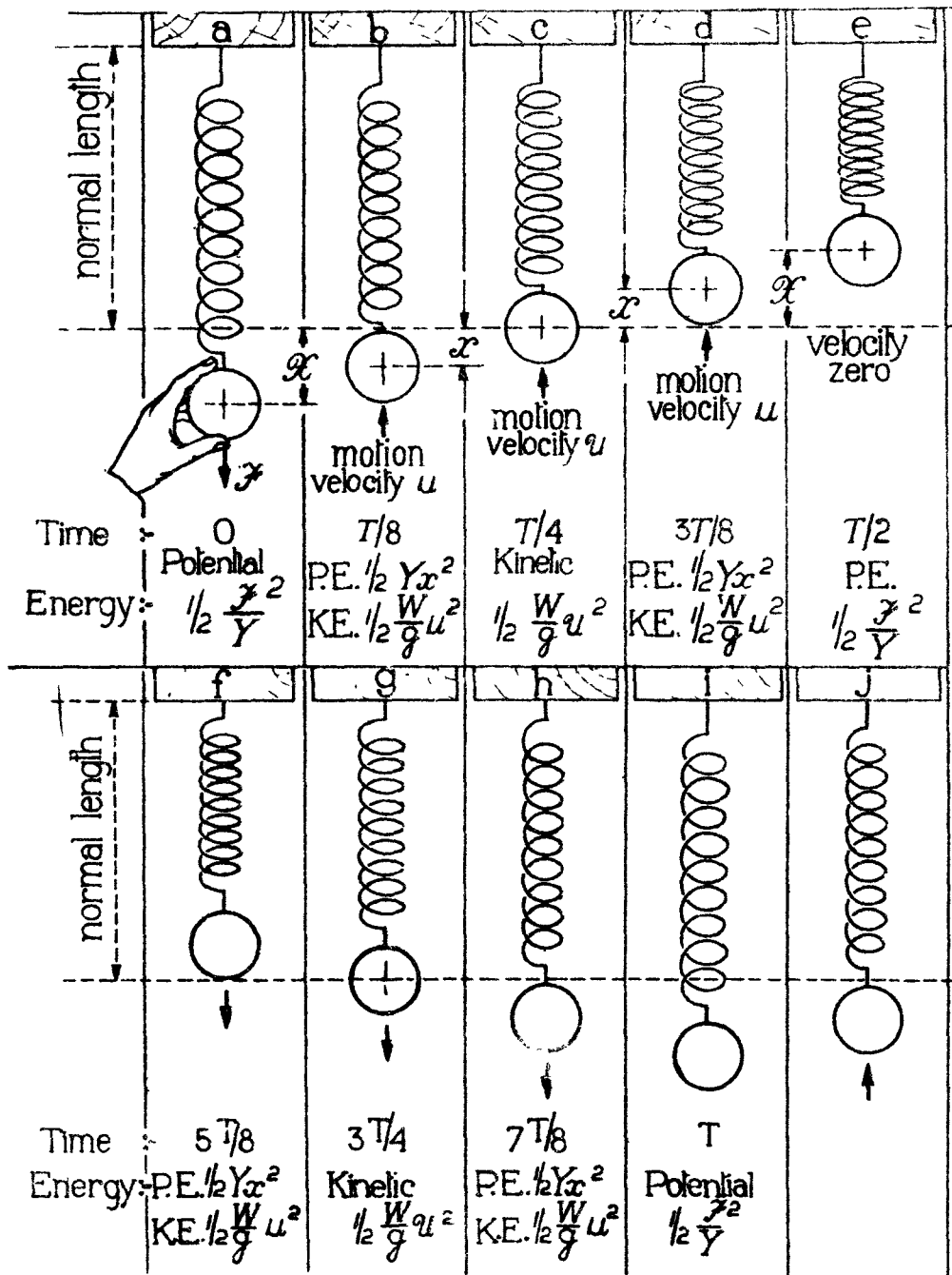
the force with which it is attracted towards the earth by gravitation. In order that the reader may himself perform the mechanical experiment to be described and may easily make calculations regarding the results, the F.P.S. (foot-pound-second) system of units will be adopted for the mechanical example, practical units being employed in the electrical analogue. The unit of mass is that upon which a force of 1 Lb. produces an acceleration of 1 ft. per second per second. Now a force of 1 Lb. produces in a mass of 1 lb. an acceleration $g = 32.2$ ft. per second per second, and if the acceleration produced by a force of 1 Lb. is only 1 ft. per second per second, the mass acted upon must be 32.2 times as great; the unit of mass is therefore 32.2 lb., and is sometimes called a "slug." In this system both kinetic and potential energy are measured in foot-pounds (ft.-Lb.) and the quantity of energy stored in a body weighing W Lb., moving with a velocity of u ft. per second, is $\frac{1}{2} \frac{W}{g} u^2$ ft.-Lb. The convention of using the symbol Lb. to denote pounds

force, and lb. to denote pounds mass, should be noted. Similarly the kinetic energy possessed by an inductance L in which a current of I amperes is flowing, is stored in the form of a magnetic field and its amount is $\frac{1}{2} LI^2$ joules. The electrical current has already been compared with the mechanical velocity, and it is apparent that the property of inductance in an electrical circuit enters into the expression for kinetic energy in the same manner as mass in the mechanical example.

5. The mechanical analogue of electrical capacitance is ability to stretch or extend, or the converse, susceptibility to compression. The latter conception has already been used to illustrate the storage of energy in a condenser by comparison with the storage of compressed air in a gas cylinder; energy is also stored when an ordinary spring is compressed or extended, for example in the mainspring of a clock. The form of spring which best lends itself to actual measurement of its property of storing energy is the spiral spring, which can be made by winding a steel wire in a screw thread, the latter being removed when the winding is completed. The amount of extension obtained from a spring of this kind is directly proportional to the force with which it is extended, provided that the force applied is insufficient to produce a permanent elongation of the spring; when the latter takes place the elastic limit is said to be exceeded. If a constant force of F Lb. produces an extension of X feet within the elastic limit, then $F \propto X$ or $F = YX$, the constant of proportion Y being a property of the particular spring in use. It is called its stiffness, and is measured by the force (in Lb.) which will produce an extension of 1 ft. In the corresponding electrical example, the quantity of electricity Q , stored in a condenser of capacitance C farads, is proportional to the applied voltage, or $V = \frac{1}{C} Q$. As the formal analogue of force is V the mechanical quantity corresponding to quantity of electricity is X , the extension of the spring, and therefore its stiffness is analogous to the reciprocal of the electrical capacitance. This signifies that a large condenser, which will acquire a given charge with only a small applied voltage, is equivalent to a weak spring, which will acquire a given extension with only a small applied force. When the condenser is charged, its P.D. being constant and equal to V volts, it possesses a supply of potential energy in the form of electrical strain in the dielectric, the quantity being $\frac{1}{2} CV^2$ joules. Similarly the stretched spring possesses a store of potential energy, equal to $\frac{1}{2} YX^2$ or $\frac{1}{2} \frac{F^2}{Y}$ ft.-Lb.

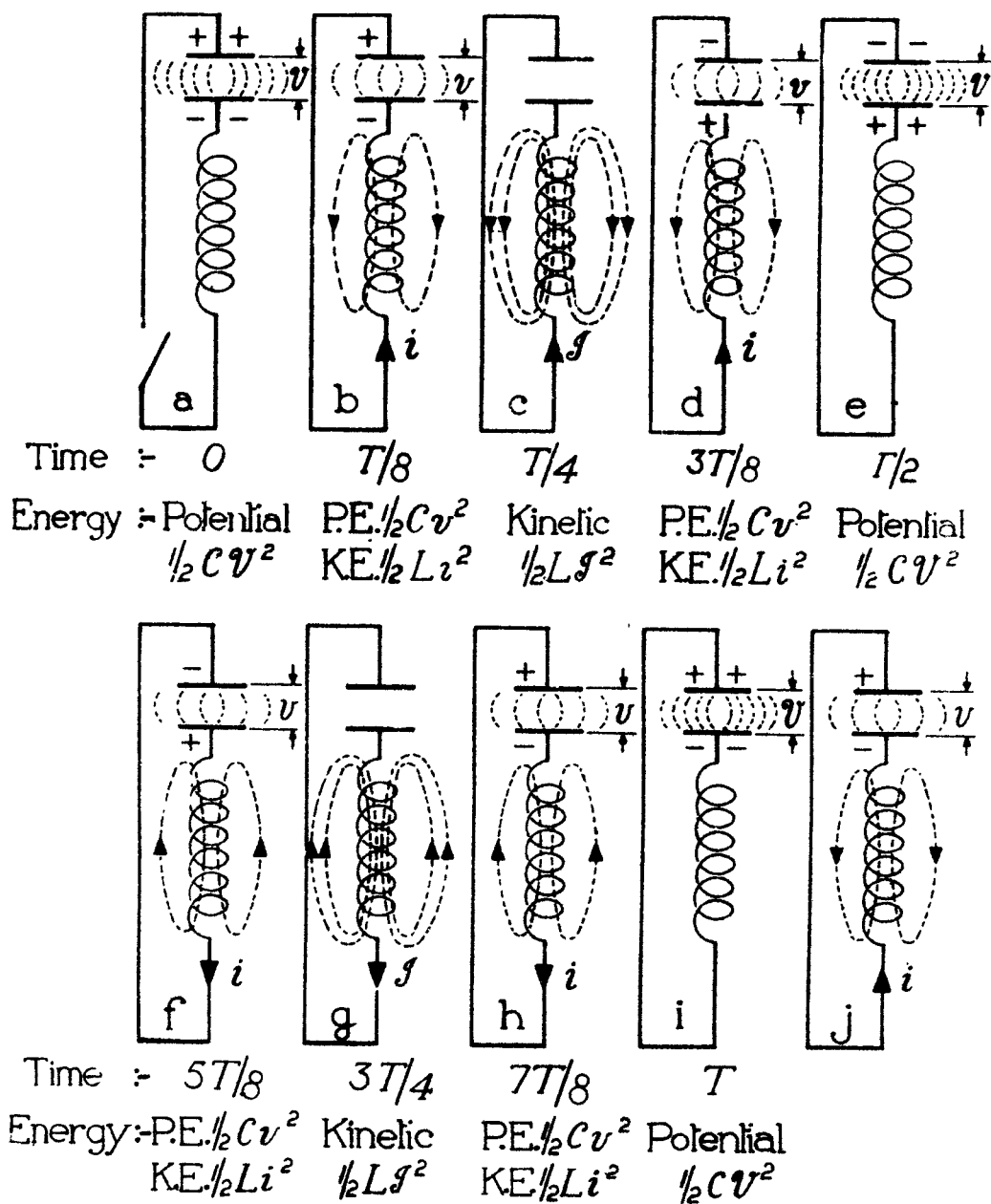
Oscillation of weighted spring

6. (i) The spring when supported at its upper end and extended by a given mass, e.g. an iron ball attached to its lower end, is said to be statically strained, and represents an electrical circuit possessing inductance, capacitance, and a very small resistance, the mechanical and electrical equivalents being shown in fig. 1. This state of static strain must be considered as the normal state of the system when the spring is arranged in the manner stated. The phenomena to be described would take place equally well if the ball and spring were arranged horizontally in such a way that the ball could slide freely along a perfectly smooth surface, but this is not practicable. The action of applying an external force to the mechanical system (fig. 2a) and



OSCILLATIONS OF WEIGHTED SPRING.

FIG.2
CHAP. VII



ELECTRICAL OSCILLATIONS

FIG 3
CHAP.VII

thus causing a further, or dynamic, extension of the spring, corresponds to the introduction of an electrical charge into a condenser (fig. 3a); in both cases, energy is stored in the system in potential form.

(ii) The action of releasing the ball corresponds to the closure of the switch S, by which the condenser is allowed to discharge through the inductance. Let us first observe the manner in which the spring loses its potential energy, neglecting the effects due to friction. Let the original applied force be $\mathcal{F}$ Lb., and the resulting displacement $\mathcal{X}$ feet. As soon as the ball is released, it is urged into motion in an upward direction by the action of the spring. Its inertia causes it to move slowly at first, but its velocity gradually increases (fig. 2b) and reaches a maximum at the exact moment when the spring has returned to its normal, i.e. statically strained, length. At this

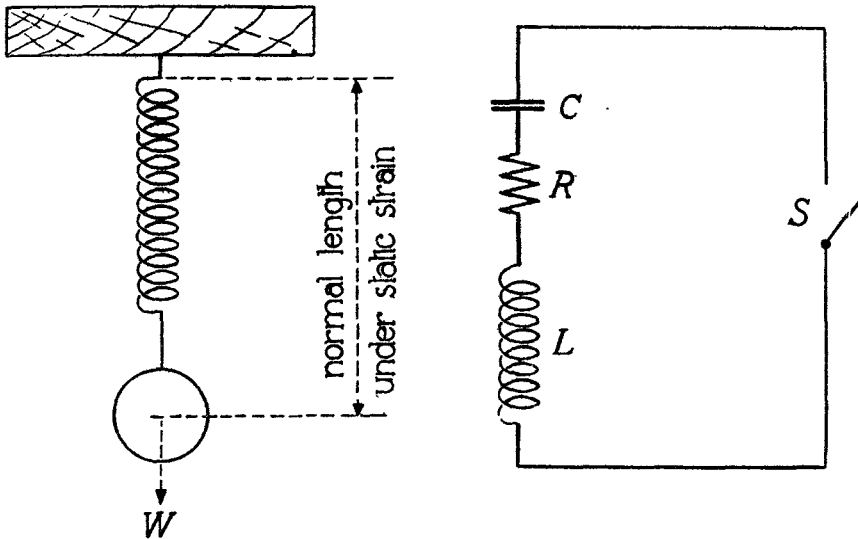


FIG. 1, CHAP. VII.—Mechanical equivalent of electrical oscillatory circuit.

moment the system possesses no potential energy, the latter having been converted into kinetic energy in the motion of the ball (fig. 2c). It is of interest to calculate the velocity of the ball at this instant. The original energy stored in the spring was $\frac{1}{2} \frac{\mathcal{F}^2}{Y}$ or $\frac{1}{2} Y \mathcal{X}^2$ ft.-Lb. and the energy stored in the moving mass is equal to the energy originally stored in the spring, hence if u is the velocity of the ball at the moment under consideration

$$\frac{1}{2} \frac{W}{g} u^2 = \frac{1}{2} \frac{\mathcal{F}^2}{Y}$$

$$u^2 = \frac{g}{WY} \mathcal{F}^2$$

$$u = \mathcal{F} \sqrt{\frac{g}{WY}}$$

A numerical example will assist in making this clear. Suppose the spring has a stiffness of 2 Lb.

CHAPTER VII.—PARA. 7

per foot, the mass of the ball to be 1 lb. and the initial extension to be 6 in. or .5 ft. Then

$$\mathcal{F} = 1 \text{ Lb.}$$

$$Y = \frac{2 \text{ Lb.}}{\text{ft.}}$$

$$\begin{aligned} u &= 1 \text{ Lb.} \times \sqrt{\frac{32 \cdot 2 \frac{\text{ft.}}{\text{sec.}}}{1 \text{ Lb.} \times 2 \text{ Lb.}}} \\ &= 4 \cdot 01 \frac{\text{ft.}}{\text{sec.}} \end{aligned}$$

(iii) After reaching its normal position the ball continues to travel upwards (fig. 2d); in doing so it does work, for it is compelled to compress the spring during its motion, and energy is required for this operation. The velocity decreases as it loses energy, and eventually the ball comes to rest $\mathcal{X}$ feet above the point at which its velocity was a maximum (fig. 2e). The energy imparted to the spring is therefore $\frac{1}{2} Y \mathcal{X}^2$ ft.-Lb. which is the amount it originally possessed, the energy being stored by compression instead of by extension. The ball is now urged downwards by the force of the spring (fig. 2f), and again attains its maximum velocity (but in the downward direction) when the spring is dynamically unstrained (fig. 2g). The velocity then decreases (fig. 2h), and finally the ball comes momentarily to rest at the bottom of its travel, the spring being again extended $\mathcal{X}$ feet; it then commences to move upward for a second time (fig. 2j), and the whole cycle shown in fig. 2 is repeated. If no energy were converted into heat and into motion of the surrounding air, the oscillation of the system would continue indefinitely. This may be expressed by saying that once a certain amount of energy has been imparted to a loss-free system of this kind the energy contained in the system remains constant, although at any instant it may be partly possessed by the spring (potential energy) and partly by the mass (kinetic energy). The sum of these two energies at any and every instant is constant and equal to the energy originally imparted, or

$$\frac{1}{2} Y x^2 + \frac{1}{2} \frac{W}{g} u^2 = \text{constant}$$

x being the displacement of the mass from its normal position, and u its velocity, at any particular instant.

Electrical oscillations

7. (i) Now consider the electrical circuit, which consists of an inductance of L henries, and a capacitance of C farads, the condenser being charged to a voltage $\mathcal{V}$ volts, so that the energy stored therein is $\frac{1}{2} C \mathcal{V}^2$ joules. On closing the switch S (fig. 3a), the condenser commences to discharge, but the growth of the current causes an increasing magnetic flux round the inductance L (fig. 3b), and consequently a counter-E.M.F. which opposes the flow of current; as a result the latter does not attain its maximum value until the moment at which the condenser is wholly discharged (fig. 3c). The energy originally stored by the dielectric in potential form is now wholly stored in the magnetic field around the inductance, and its amount is $\frac{1}{2} L \mathcal{J}^2$ joules. Knowing the amount of energy originally stored, we can find the current at this instant, from the

known conditions that $\frac{1}{2} L \mathcal{J}^2 = \frac{1}{2} C \mathcal{V}^2$. Solving this equation for $\mathcal{J}$ it is found that $\mathcal{J} = \mathcal{V} \sqrt{\frac{C}{L}}$.

We may compare this with the expression obtained for the velocity of the mass in the mechanical system, i.e. $u = \mathcal{F} \sqrt{\frac{g}{Y W}}$. The mass $\frac{W}{g}$ and inductance L occupy similar positions in the formulae, while $\frac{1}{Y}$ and C also correspond.

(ii) Owing to the counter-E.M.F. of self-induction caused by the changing magnetic flux, the current will continue to flow, charging the condenser with a polarity opposite to its original charge (fig. 3d). As soon as any charge is introduced into the condenser, it exerts a counter-E.M.F. which opposes the introduction of an additional quantity of electricity and therefore the current decreases with a consequent collapse of magnetic flux round the inductance. The change of magnetic flux in turn sets up a forward E.M.F. which tends to maintain the current against the back pressure of the condenser, but eventually the current falls to zero, together with the magnetic flux, and the whole of the energy which was stored in the inductance when the current was a maximum, is now again stored in the condenser (fig. 3e), its quantity being $\frac{1}{2} C V^2$ joules as originally. The condenser will now discharge through the inductance once more (fig. 3f), the current flow being in the opposite direction. At the moment when the condenser is completely discharged (fig. 3g), the current in the circuit will again be a maximum, and the counter-E.M.F. caused by the collapse of the flux will cause the condenser to charge with its original polarity (fig. 3h). When this is accomplished (fig. 3i) one complete cycle of oscillation has been performed. The characteristic property of the oscillators we have described, both mechanical and electrical, is that at any instant the total energy possessed by the circuit is constant and equal to that originally imparted since the effects of friction in the mechanical system and resistance in the electrical circuit have both been neglected.

Example 1.

A condenser of .01 microfarad is given a charge of Q of 20×10^{-6} coulomb, and its plates then connected by a coil of 100 microhenries having negligible losses. Find (a) the initial voltage of the condenser, (b) the maximum current during discharge, (c) the total energy, W , stored in the circuit and (d) the current at the instant when the condenser P.D. is 1,500 volts

Since $Q = CV$

$$V = \frac{Q}{C} = \frac{20 \times 10^{-6}}{.01 \times 10^{-6}}$$

$$= 2,000 \text{ volts}$$

$$I = V \sqrt{\frac{C}{L}}$$

$$= 2,000 \sqrt{\frac{.01}{100}}$$

$$= 20 \text{ amperes}$$

$$W = \frac{1}{2} C V^2 \text{ joules.}$$

$$= \frac{1}{2} \times .01 \times 10^{-6} \times (2 \times 10^3)^2$$

$$= \frac{1}{2} \times .01 \times 4$$

$$= .02 \text{ joules.}$$

The energy w stored in the condenser when $v = 1,500$ volts, is $\frac{1}{2} C v^2$ joules.

$$w = \frac{1}{2} \times .01 \times 10^{-6} \times (1.5 \times 10^3)^2$$

$$= \frac{1}{2} \times .01 \times 2.25$$

$$= .01125 \text{ joules.}$$

The energy stored in the inductance is .02—01125 or .00875 joules, and is equal to $\frac{1}{2} L i^2$ joules.

$$L = 100 \times 10^{-6} \text{ henry}$$

$$\frac{1}{2} \times 100 \times 10^{-6} \times i^2 = .00875$$

$$10^{-4} i^2 = .0175$$

$$i^2 = 175$$

$$i = 13.2 \text{ amperes.}$$

CHAPTER VII.—PARAS. 8-9

Period and frequency of free oscillation

8. (i) The reader is strongly advised to perform the above experiment with a spiral spring and weight for himself. A spring which has been found very suitable for the purpose may be obtained from an old roller blind, this being generally of steel, about fifteen inches long when not extended, and having a stiffness of about 5 Lb. per ft. It will carry about double this weight without exceeding the elastic limit. If the duration of one complete oscillation is measured, it will be found to be given approximately by the formula $T = 2\pi \sqrt{\frac{W}{gY}}$, this time being called the period of oscillation. The stiffness Y can be measured by noting the initial extension given by the mass actually used for the experiment. If the stiffness is 5 Lb. per ft. and the weight of the ball is 2 Lb., the time of one complete oscillation will be approximately 0.7 second. In the electrical circuit which we have stated to be exactly analogous to the above, the duration of one complete oscillation can be found by substituting L , the value of the inductance in henries for the mass $\frac{W}{g}$, and C , the value of the capacitance in farads, for the reciprocal of the stiffness

$$\left(\frac{1}{Y}\right), \text{ giving } T = 2\pi \sqrt{LC}.$$

(ii) The period of an electric oscillatory circuit can actually be measured, just as the period of a mechanical oscillation can, and this equation is found to be approximately true. It is not entirely so owing to the effect of friction in the mechanical oscillator and resistance in the electrical one. The true expression in the electric case is

$$T = 2\pi \frac{1}{\sqrt{\frac{1}{LC} - \frac{R^2}{4L^2}}}.$$

The frequency of the oscillation is the number of complete cycles executed in one second, and obviously if T is the duration of one cycle of oscillation, there are $\frac{1}{T}$ cycles per second. The

frequency being denoted by f_n , $f_n = \frac{1}{2\pi} \sqrt{\frac{1}{LC} - \frac{R^2}{4L^2}}.$

(iii) When any system, whether mechanical or electrical, is set into oscillation by an initial stress or charge, and the frequency depends entirely upon the mass, stiffness and friction, or upon the inductance, capacitance and resistance, it is said to be in a state of free oscillation, and the frequency f_n at which it freely oscillates is said to be its natural frequency. Another type of oscillation may take place, for example, when a force of given frequency is applied to the mechanical system, forced oscillations take place at this frequency. The electrical parallel of this is the alternating current which is established when an E.M.F. of any frequency whatever is applied to an electrical circuit; this type of oscillation has been dealt with in Chapter V.

9. So far we have not considered the manner in which the displacement of the vibrating body varies with time. The motion of the ball undergoes the following variations. Starting at the bottom of its available travel, it commences to move upwards, and reaches its maximum velocity in the middle of its path, then receives negative acceleration because it is doing work (i.e. losing energy) in compressing the spring and temporarily comes to rest with maximum upward displacement. It then receives downward acceleration, possessing maximum velocity in the midpoint of its travel, and again reaches a position of momentary rest at its normal (i.e. statically strained) position. If the displacement of the ball in this path is plotted at every instant over a number of complete cycles it will be seen that the graph connecting displacement with time is of sinusoidal form; since the displacement has its maximum value at the time $t = 0$, the graph is actually a cosine curve, and the displacement x , after an interval of t seconds is

$$x = X \cos \omega t.$$

In the electrical circuit, the P.D. between the condenser plates also follows a cosine law, i.e. $v = V \cos \omega t$, but the current in the circuit is zero when the condenser is initially charged, and grows in value as the condenser P.D. falls, attaining its maximum value when the condenser P.D. is zero. The current continues to flow in the same direction, charging the condenser with reverse polarity. Thus the current variation obeys the law $i = I \sin \omega t$; the relative phase between the condenser P.D. and current is therefore as shown in fig. 4, the loss of energy being assumed to be negligible. If the resistance losses are taken into account, the displacement curve will still follow the cosine law, but its amplitude will diminish in every succeeding half-cycle, signifying that a certain portion of the energy is converted into heat or some other unrecoverable form during every oscillation.

10. The acceleration of the vibrating mass may also be considered. We have seen that when the displacement is a maximum in the downward direction, the acceleration is upward and of maximum value, gradually decreasing in value as the velocity increases, and becoming zero

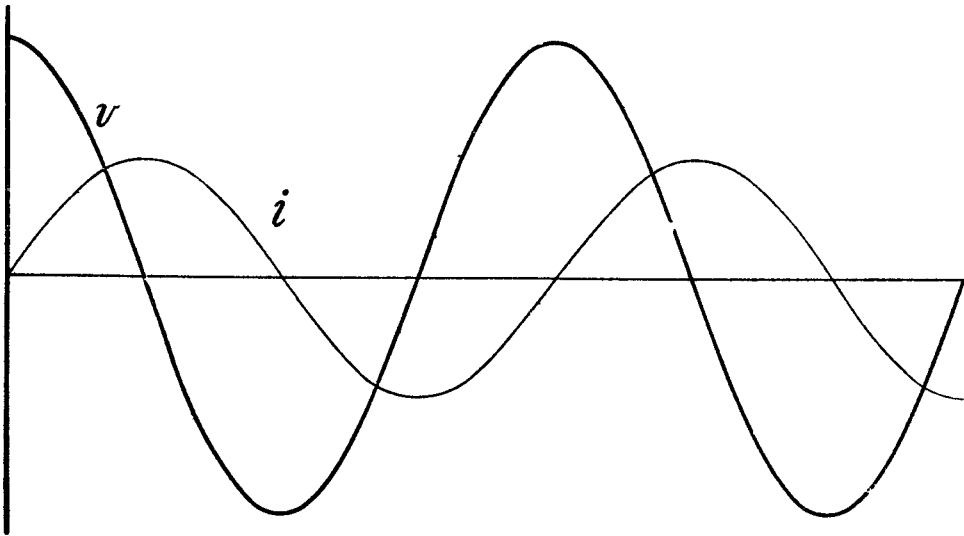


FIG. 4, CHAP. VII.—Relative phase of oscillatory current and condenser P.D.

when the velocity is greatest. The acceleration then becomes negative, reducing the velocity, and reaching a maximum negative value when the ball is at its highest point and its velocity zero, so that at every instant the acceleration is of the opposite sign to the displacement; if plotted on the same time scale as the displacement and velocity, the acceleration is seen to obey a cosine law, but is of opposite sign to the displacement. This relation between the phases of displacement, velocity and acceleration is a characteristic always possessed by a body or system which is executing the type of motion under discussion, i.e. simple harmonic motion. This is nothing more than another illustration of the law formulated in Chapter V for the rate of change of a sinusoidal quantity, i.e. if $x = X \cos \omega t$, where x is the instantaneous and X the maximum displacement, and $\omega = 2\pi f$,

$$\text{the velocity, } \frac{dx}{dt} = -\omega X \sin \omega t$$

$$\begin{aligned} \text{and the acceleration, } \frac{d^2 x}{dt^2} &= -\omega^2 X \cos \omega t \\ &= -\omega^2 x. \end{aligned}$$

CHAPTER VII.—PARA. 11

Effect of resistance upon the natural frequency

11. (i) It has been stated that in an electrical circuit having such low resistance that its influence upon the period is negligible, the period of one complete oscillation is calculated from the equation $T = 2\pi\sqrt{LC}$, L being expressed in henries and C in farads. If the resistance is large, however, it may be necessary to take its value into consideration in calculating the period, and it becomes very desirable to appreciate the conditions under which it is permissible to use the approximate formula, and when it is necessary to employ the exact expression $T = \frac{2\pi}{\sqrt{\frac{1}{LC} - \frac{R^2}{4L^2}}}$.

It is perhaps more convenient to deal with frequency instead of period, and the formulæ under discussion are then

$$f_n = \frac{1}{2\pi} \sqrt{\frac{1}{LC}} \text{ approximately} \quad \dots \dots \dots (a)$$

$$f_n = \frac{1}{2\pi} \sqrt{\frac{1}{LC} - \frac{R^2}{4L^2}} \quad \dots \dots \dots (b),$$

which is rigorously true if the resistance term includes all sources of loss of energy. In order to appreciate the error which is introduced by using formula (a) let us calculate the natural frequency of a typical oscillatory circuit consisting of a coil of $200 \mu H$ inductance connected to a capacitance of $0.0002 \mu F$, the resistance being variable. The natural frequency of the circuit neglecting the effect of resistance, is given by formula (a).

$$L = \frac{200}{10^6} \text{ henries} = \frac{2}{10^4} \text{ henries}$$

$$C = \frac{0.0002}{10^6} \text{ farads} = \frac{2}{10^{10}} \text{ farads}$$

$$LC = \frac{4}{10^{14}} \text{ and } \sqrt{LC} = \frac{2}{10^7}$$

$$f_n = \frac{1}{2\pi} \times \frac{10^7}{2} = \frac{1}{\pi} \times 2.5 \times 10^6$$

$$\text{Now } \frac{1}{\pi} = 0.31831$$

$$\therefore f_n = 0.31831 \times 2.5 \times 10^6$$

$$= 795775 \text{ cycles per second.}$$

(ii) Now let it be assumed that the total resistance of the circuit is 40 ohms, and allow for this in calculating the natural frequency by formula (b).

The value of $\frac{1}{LC}$ is already known to be $\frac{10^{14}}{4}$ and we proceed to calculate the value of $\frac{R^2}{4L^2}$.

$$\frac{R}{2L} = \frac{40 \times 10^6}{2 \times 200} = 10^5, \quad \frac{R^2}{4L^2} = 10^{10}.$$

From formula (b) therefore,

$$\begin{aligned} f_n &= \frac{1}{2\pi} \sqrt{\frac{10^{14}}{4} - 10^{10}} \\ &= \frac{1}{2\pi} \sqrt{2500 \times 10^{10} - 1 \times 10^{10}} \\ &= \frac{10^5}{2\pi} \sqrt{2499} \end{aligned}$$

instead of $\frac{10^5}{2\pi} \sqrt{2500}$ which is the natural frequency if the resistance is zero. The error introduced by neglecting the resistance and using the approximate formula is in this instance only about 0.02 per cent. It may be taken as a general rule that no useful purpose will be served by calculating the frequency to a greater degree of arithmetical accuracy than the accuracy with which the values of inductance, capacitance and resistance are known. It was, of course, assumed above that the values of L and C were precisely those given, but in practice it is unlikely that either L or C will be known within one or two per cent. of the true values hence the labour expended in calculating the natural frequency by means of formula (b) is not repaid in the form of a more accurate answer.

12. (i) In certain circumstances, resistance may be deliberately introduced into the circuit, and it may then become necessary to calculate the natural frequency by the exact formula. Thus, if a total resistance of 707 ohms exists in a circuit having the above inductance and capacitance,

$$\begin{aligned}
 f_n &= \frac{1}{2\pi} \sqrt{\frac{1}{LC} - \frac{R^2}{4L^2}}, \\
 \frac{R}{2L} &= \frac{707}{2} \times \frac{10^6}{200}, \text{ and as } 707 = \frac{1000}{\sqrt{2}} \\
 \frac{R}{2L} &= \frac{10^7}{4\sqrt{2}}, \quad \frac{R^2}{4L^2} = \frac{10^{14}}{32} = 3.12 \times 10^{12} \\
 \frac{1}{LC} &\text{ remains as before, viz., } 25 \times 10^{12}. \\
 \therefore \frac{1}{LC} - \frac{R^2}{4L^2} &= (25 - 3.12) \times 10^{12} \\
 &= 21.88 \times 10^{12} \\
 \text{and } f_n &= \frac{10^6}{2\pi} \sqrt{21.88} \\
 &= \frac{10^6}{2\pi} \times 4.6776 \\
 &= 744,470 \text{ cycles per second.}
 \end{aligned}$$

The natural frequency has been reduced by approximately 6 per cent. owing to the insertion of this comparatively large resistance.

(ii) Now suppose the resistance to be increased still further, say to 1,414 ohms. A repetition of the calculation then gives the natural frequency as 562,600, i.e. a reduction of about 30 per cent. As the resistance added to the circuit is still further increased, the natural frequency decreases rapidly, and if the total resistance of the circuit becomes 2,000 ohms, it is found that $\frac{R^2}{4L^2} = \frac{1}{LC}$, the quantity beneath the square root sign in formula (b) becomes zero, and the arithmetical value of the natural frequency also zero. It is important to understand the physical meaning of this, and here the mechanical analogy greatly assists. Suppose that the ball which is suspended on the spring, and possibly the spring itself, is immersed in a viscous fluid, e.g. treacle. When motion is taking place, the friction will be much greater than when the motion occurs in air, and in the extreme instance, it may be found that on pulling down the ball, it merely returns to its original position without oscillation, because the work done in overcoming the friction during the first upward motion is equal to the potential energy originally stored in the spring. The mass then possesses no kinetic energy at the instant when it reaches its normal position, and is therefore unable to do work on the spring by compressing it. The whole

CHAPTER VII.—PARA. 13

mechanical system, although it possesses both inertia and elasticity, is now non-oscillatory, because its friction is excessive. In an electric circuit possessing both capacitance and inductance, the introduction of a sufficient amount of resistance also renders the system non-oscillatory.

Damping

13. (i) The term $\frac{R}{2L}$, in the expression for the natural frequency of an electrical circuit, is called the damping factor. When the damping factor is exactly sufficient to prevent oscillation, i.e. when $\frac{1}{LC} = \frac{R^2}{4L^2}$ the circuit is said to be critically damped; the amount of resistance necessary to ensure this for any given values of L and C is called the critical resistance, and is derived thus

$$\begin{aligned}\frac{1}{LC} &= \frac{R_c^2}{4L^2}, \quad R_c \text{ being the critical value of } R. \\ R_c^2 &= \frac{4L^2}{LC} \\ &= \frac{4L}{C} \\ \therefore R_c &= 2\sqrt{\frac{L}{C}}\end{aligned}$$

Any value of resistance greater than this will cause the condenser discharge to be uni-directional, no oscillation being produced, and the larger the resistance, the longer will the condenser take to become fully discharged. A circuit which possesses an amount of total effective resistance which is sufficient or more than sufficient to prevent free oscillation is said to be aperiodic.

(ii) Although the influence of a small resistance upon the natural frequency of a circuit is negligible, the presence of resistance in an oscillatory circuit has another effect which is of the greatest importance. If set into oscillation, an electrical circuit without resistance, or a mechanical circuit without friction, would continue to oscillate with undiminishing amplitude for ever, for its total store of energy would never be depleted, although it would sometimes be in kinetic and sometimes in potential form, and during a large portion of each period it would be partly in one form and partly in the other. In practice a certain amount of energy is converted into heat and into motion of air in the mechanical system, while in the electrical circuit energy is converted into heat and is also expended in other ways to be discussed later. The energy stored at the end

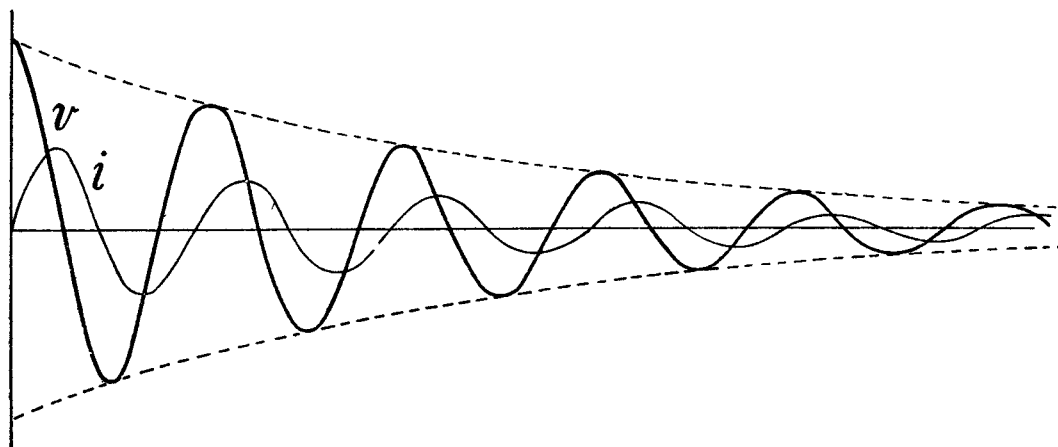


FIG. 5, CHAP. VII.—Damped oscillation. Current and condenser P.D.

of each succeeding half cycle becomes progressively less, the amount wasted during each cycle being a certain percentage of that which was stored at the commencement of the cycle. If the original energy is considered to be 100 per cent. and 20 per cent. is wasted in the first complete cycle, 80 per cent. of the energy is available for the second cycle, and during the second cycle 20 per cent. of 80 per cent., i.e. 16 per cent., of the original energy is lost, the energy remaining being 64 per cent. Both the peak value of the condenser P.D. and the peak value of the oscillatory current will decrease in amplitude during each successive half-cycle, and it is more usual to express the damping in terms of either of these amplitudes rather than in terms of total energy. Fig. 5 shows the manner in which the amplitudes of voltage and current decay with time, and it will be observed that each successive peak touches a curve (shown by the dotted line) which is similar in shape to the curve showing the discharge of a condenser through a resistance only (fig. 25, Chapter I).

Logarithmic decrement

14. The ratio of the amplitude of one peak to that of the same sign which follows it is called the decrement of the oscillation. If the successive amplitudes of positive condenser P.D. are 100, 90, 81, 72·9, 65·6, etc., the decrement is $\frac{100}{90} = \frac{90}{81} = \frac{81}{72·9}$ etc. or 1.11..... It is more convenient however to use a quantity called the logarithmic decrement (which is the naperian logarithm of the decrement), the naperian logarithm being 2·3026 times the ordinary or common logarithm generally found in mathematical tables. This expression is often written and referred to as the log. dec., and is related to the magnification of the circuit. The latter term has been explained in Chapter V, and it will be recalled that the magnification χ is given by the ratio $\frac{\omega_r L}{R} = \frac{1}{\omega_r C R}$ where ω_r is the resonant frequency of the circuit, or the frequency at which an applied E.M.F. of given amplitude will cause maximum current to flow. The log. dec. may be derived from consideration of the oscillatory properties of the circuit by observing that it is the ratio of the energy expended in various losses during any one half-cycle to the maximum energy held in either kinetic or potential form during the same time. If the peak value of the current of a given half-cycle is $\mathcal{I}$, the heating effect during the half-cycle is $\frac{\mathcal{I}^2 R}{2} \times \frac{T}{2}$ joules, R being the total effective resistance and $\frac{T}{2}$ the duration of the half-cycle. The amount of energy stored in the circuit in kinetic form at the instant when the current reaches peak value is $\frac{1}{2} L \mathcal{I}^2$ joules, and the ratio referred to above becomes, using the usual symbol δ to denote the log. dec.,

$$\begin{aligned}\delta &= \frac{\text{Energy wasted}}{\text{Energy stored}} = \frac{\frac{\mathcal{I}^2 R}{2} \times \frac{T}{2}}{\frac{1}{2} L \mathcal{I}^2} \\ &= \frac{RT}{2L}, \text{ but since } T = \frac{1}{f_n}, \\ \delta &= \frac{R}{2Lf_n}\end{aligned}$$

15. In circuits which are designed for the production of oscillations, i.e. in which resistance has not been deliberately introduced, the frequency f_n may be replaced with negligible error by the resonant frequency $f_r = \frac{1}{2\pi\sqrt{LC}}$ and the log. dec. becomes

$$\begin{aligned}\delta &= \frac{R}{2L} \times 2\pi\sqrt{LC} \\ &= \pi R \sqrt{\frac{C}{L}}\end{aligned}$$

CHAPTER VII.—PARA. 15

It has been previously shown that the magnification of the circuit is given by the expression

$$x = \frac{\omega_r L}{R} = \frac{1}{R} \sqrt{\frac{L}{C}}$$

It is immediately apparent that $\delta = \frac{\pi}{x}$ and the importance of this relation lies in the fact that it is comparatively easy to measure the magnification of a circuit, at any rate up to frequencies of the order of a million cycles per second, and hence to calculate the log. dec. For example, if the magnification of a given circuit is found by actual measurement to be 157, the log. dec. is $\frac{\pi}{157} = \frac{3.14}{157} = .02$, and this is the naperian logarithm of the ratio $\frac{I_1}{I_2}$. The common log of $\frac{I_1}{I_2}$ is therefore $.02 \times .4343$ or $.008686$, and reference to a table of common logs shows that the corresponding natural number is 1.0202. Hence $I_1 = 1.0202 I_2$. For some purposes it is more convenient to refer to the ratio $\frac{I_2}{I_1}$ which in this instance is $.9802$, and $I_2 = .9802 I_1$. This inverted form of the decrement is called the persistency of the oscillation, and is conveniently expressed as a percentage of the initial amplitude. In the above example the persistency would be described as 98.02 per cent.

Example 2

In fig. 5 suppose the initial condenser P.D. to be 200 volts, and the amplitude of the next positive peak to be 131.2 volts. Find the log. dec. of the circuit. If $L = 100$ microhenries and $C = .001$ microfarad, find also the resistance of the circuit and its magnification.

The decrement is $\frac{200}{131.2} = 1.524$

Common log. of 1.524 = .1830

Naperian log. of 1.524 = $\delta = .1830 \times 2.3026$

$$\delta = .4214$$

Since

$$\delta = \pi R \sqrt{\frac{C}{L}}$$

$$R = \frac{\delta}{\pi} \sqrt{\frac{L}{C}}$$

$$= \frac{.4214}{3.1416} \sqrt{\frac{100}{.001}}$$

$$= \frac{.4214}{3.1416} \sqrt{100,000}$$

$$= .4214 \times \frac{316}{3.1416}$$

$$= 42.2 \text{ ohms}$$

and

$$x = \frac{\pi}{\delta}$$

$$= \frac{3.1416}{.4214}$$

$$= 7.44$$

16. (i) The effect of resistance upon the oscillatory properties of a circuit is shown graphically in fig. 6. The oscillatory current in a perfectly loss-free circuit is shown by the sine curve of constant amplitude, while the current in a lightly damped circuit will have a slightly greater period as in the second curve. The third curve shows the growth and decay of the current when $R = 2\sqrt{\frac{L}{C}}$, i.e. when the circuit is critically damped ; it will be seen that the current reaches its maximum value rather more rapidly than in an oscillatory circuit, but then dies away without reversal of sign. Finally, the growth and decay of current in a heavily damped circuit is shown. The current rises more rapidly and dies away more slowly than in a critically damped circuit. In discussing mechanical and electrical oscillations the decay of the mechanical oscillation of the spring and ball was ascribed to friction, and that of the electrical oscillation to resistance, without any consideration of the forms such friction and resistance may take. In the mechanical

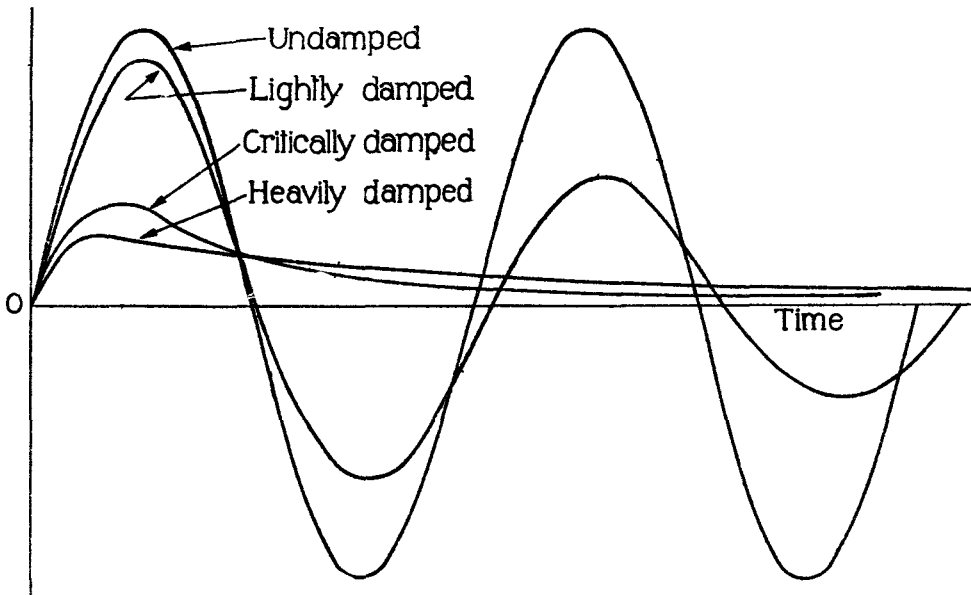


FIG. 6, CHAP. VII.—Effect of resistance upon condenser discharge.

system, the vibration is communicated to the surrounding atmosphere, and if the frequency of vibration is sufficiently high, results in the production of a sound wave in the air. The energy borne by this sound is, of course, supplied by the oscillatory system, and constitutes one form of damping. In the electrical circuit, the causes of energy loss may be divided into five components ; these are :—

- (i) Conductor resistance.
- (ii) Eddy currents in neighbouring conductors.
- (iii) Condenser losses in the actual electric circuit.
- (iv) Dielectric losses in the surrounding insulators.
- (v) Hysteresis losses, if iron is present.
- (vi) Radiation of energy in the form of electro-magnetic waves.

(ii) The oscillatory circuit which has been discussed hitherto consists of a condenser, inductance and resistance in series, the condenser being tacitly assumed to be of the parallel plate type. Such a circuit is known as a closed oscillatory circuit in contrast with another type

CHAPTER VII.—PARAS. 17-19

which is called an open oscillatory circuit. In this form the condenser plates are opened out as much as is practicable, in order that the greatest possible volume of dielectric may be embraced by the electric field of the condenser. Under these conditions the amount of energy converted into electro-magnetic radiation is much greater than in a closed oscillator. Before further discussion of the properties of open oscillators, it is necessary to explain what is meant by an electromagnetic wave.

ELECTROMAGNETIC WAVES

Wave motion

17. A disturbance which is propagated in any medium in such a manner that the shape, but not necessarily the magnitude, of the disturbance is repeated at regular intervals in space and time is called a wave, and the passage of such a disturbance through the medium is termed wave motion. A familiar example of wave motion is the surface wave caused by dropping a stone vertically into a pond. The water displaced downwards by the impact rises to occupy its former position, but owing to its inertia actually rises a little higher, forming momentarily a small mound of water upon the surface. This condition is evidently an unstable one, and the displaced volume of water immediately descends, passing its original position once more, continuing to oscillate up and down with decreasing amplitude until the energy imparted to the water by the falling stone has all been dissipated. The oscillation is communicated to the surrounding water in the immediate vicinity of the disturbance, and a wave spreads outward in ever increasing circles. The kinetic energy of the stone has been partly converted into wave motion, instead of being wholly dissipated in producing heat, as would have been the case if its fall had been arrested by a perfectly rigid earth. It must be fully realised that in no case of wave motion do the particles of the undulating medium travel forward with the wave but only move up and down, or to and fro, through a limited path. An illustration of this is found in the motion of a cork floating in the pond in which a surface wave has been produced, the cork moving up and down as the water beneath it becomes alternately the crest or trough of a wave, but possessing no average velocity along the surface of the water, although the energy imparted to the water is carried outward in ever widening circles.

Properties of waves

18. A wave can be described by means of four characteristic properties, (i) its velocity of propagation, (ii) its frequency, (iii) its amplitude and (iv) its wavelength. The *velocity of propagation* is the velocity with which the energy of the wave is conveyed from point to point in the medium, the *frequency* being the number of complete cycles of disturbance which pass a given point in unit time, i.e. one second. A *cycle* is one complete series of variations of displacement between adjacent repetitions of the wave in space, the *amplitude* or *peak value* is the maximum displacement of the medium from its normal position, and the *wavelength* is the distance between corresponding states of displacement in two adjacent repetitions of the wave form. It may be visualised as the distance between two adjacent crests or troughs. The most elementary form of wave motion is that in which each particle of the medium performs a cycle of displacement with simple harmonic motion, that is a "to and fro" or "up and down" reciprocating motion. When sinusoidal wave motion is taking place in a material medium each particle carries out this simple harmonic motion in succession. The term transverse wave motion is applied to a wave in which the particles execute their motion in a direction perpendicular to the direction of propagation of the wave, as for example, the water particles in the surface wave mentioned above. In many instances the particles vibrate to and fro about their mean position in the direction of propagation, and this mode of vibration gives rise to a longitudinal wave. The vibration of air particles when conveying sound energy is executed in this manner.

19. The relation between the frequency, velocity of propagation and wavelength of a wave in any medium depends upon the physical properties of the medium, sometimes in a complex manner because in some media the velocity of propagation is itself dependent upon the frequency. Such a medium is said to be dispersive, while a medium in which the velocity of propagation is

constant for all frequencies is called a non-dispersive medium. In non-dispersive media, to which the present discussion will be confined, there is a simple relation between the wavelength λ , velocity of propagation, u , and frequency f . Since one wavelength is the distance travelled in the periodic time, T , i.e. the time taken to execute one cycle of vibration,

$$\lambda = uT$$

and as T and f are in reciprocal relationship

$$\lambda = \frac{u}{f}$$

if u is in metres per second, λ is in metres. The mathematical expression for a sinusoidal wave moving through a medium with velocity u is

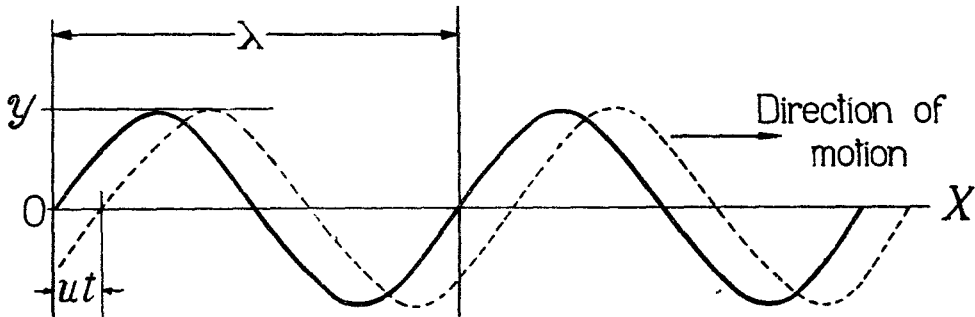
$$y = \mathcal{Y} \sin \frac{2\pi}{\lambda} (x - ut)$$

y being the displacement of a particle at time t , x the distance of the particle from the origin, and $\mathcal{Y}$ the maximum displacement of the particle. This equation may be deduced thus:—The expression for the motion of the particles about their mean position is

$$y = \mathcal{Y} \sin \omega t$$

whereas the equation of a sine wave, stationary in space on the axis OX (fig. 7) is

$$y = \mathcal{Y} \sin kx$$



Full line shows wave at instant when $t = 0$

Dotted line shows wave after an interval t

FIG. 7, CHAP. VII.—Wave travelling with velocity u .

k being a constant, having dimensions "angle per unit distance" so that kx represents an angle although x represents distance. If the wave form moves forward in space with velocity u the distance passed through in t seconds is ut , and therefore

$$y = \mathcal{Y} \sin k(x - ut).$$

The value of the constant k can now be found; when the wave has moved forward one wavelength λ , the angle moved through is 2π radians and therefore $k\lambda = 2\pi$, or $k = \frac{2\pi}{\lambda}$. The complete equation is therefore as stated above.

The ether

20. In an earlier paragraph an allusion was made to the motion of a cork upon a water surface consequent upon the impact of a stone at some other point in the wave-supporting medium. The wave does work on the cork, and so the stone may be considered as the original

CHAPTER VII.—PARAS. 21-22

source of energy and the cork as a receiver of energy. In the solar system, energy is conveyed from point to point in this manner without the intervention of any material medium whatever, the term material medium signifying one which consists of matter in either of the three forms, solid, liquid or gas. For instance large quantities of energy are received by the earth from the sun, and we are conscious of its reception by the sensations of heat and light, although the space between the sun and the earth is to all intents and purposes devoid of matter. It is possible to prove that this energy is conveyed by wave motion, and the question arises, by what medium is the energy conveyed?

21. It must be admitted that no perfectly satisfactory answer to this question has yet been evolved, and it is therefore necessary to assume that some medium exists which is capable of and responsible for the conveyance of energy from point to point in the universe. To this hypothetical medium the name "luminiferous (i.e. light-bearing) ether" was originally given, nowadays generally shortened to "the ether". A further essential assumption is that the ether penetrates all matter and pervades all space, so that it fills impartially the space between electrons and protons in the atom as well as between planets in the solar system and between stars in the universe. It is possible to prove by astronomical observation as well as by direct terrestrial experiment that wave motion, e.g. light, is propagated through this medium with the enormous velocity of 2.998×10^{10} centimetres per second (generally taken as 3×10^{10} centimetres per second). In order to appreciate the significance of this statement, this velocity should be compared with that of a compression wave in hard steel, i.e. approximately 5.2×10^5 centimetres per second, which is almost the maximum velocity with which wave motion can be conveyed by matter. Sound waves in air travel much more slowly, a mere 3.4×10^4 centimetres per second. Now it can be proved that in order to support wave motion, a material medium must possess two properties; first, a constraint tending to restore the displaced particles of the medium when the displacing force or stress is removed, which is called the elasticity of the medium, and second, inertia, the tendency of any element of the medium to continue in a state of rest, or of uniform motion in a straight line. The inertia of the particles causes them to pass through their original positions when returning thereto under the influence of the restoring force or elasticity.

22. The velocity u of wave motion in a material medium of density d and elasticity η is given by the equation $u = \sqrt{\frac{\eta}{d}}$. Since the ether has been postulated as the medium of wave motion in the conveyance of energy in the form of light and heat, it must be assumed to possess properties which are analogous to elasticity and density. It is neither necessary nor desirable to speculate upon the exact nature of these properties, but it is known that the magnetic permeability (μ) of the medium is connected with the etherial analogue of inertia and the dielectric constant or permittivity (κ) with the property corresponding to elasticity. An attempt will therefore be made to exhibit the manner in which these two properties are involved in the velocity of propagation of an electro-magnetic wave. We have already suggested (Chapter I) that an electric field of strength E may be considered to represent electrical stress, the corresponding strain being the electric flux density D ; the two quantities are related by the equation $E = \frac{4\pi D}{\kappa}$. Similarly

the magnetic field strength H and the resulting flux density B are related by the equation $H = \frac{B}{\mu}$.

The occurrence of the constant 4π in the relation between the electric quantities but not in the corresponding magnetic equation, may be regarded as the result of an unfortunate departure from uniformity in the definitions of unit electric charge and unit magnetic pole.

Production of electric field by magnetic field in motion

23. Faraday's law may be taken as a convenient starting point in this discussion. This states that in an electric circuit, an induced E.M.F. results from any and every change in the flux linkage. The point to bear in mind here is that the word circuit does not imply that the current path is conductive throughout, for it may contain an electrical condenser. Once this

point is realised, there is no necessity for the circuit to contain any conductor whatever, a change of flux through a small area of dielectric, for example, will set up an electro-motive force round the line bounding that area. In Chapter II it was stated that a magnetic flux of density B (E.M. units) perpendicularly cutting a conductor with velocity u centimetres per second, generates in it an E.M.F. of Bu electromagnetic units per centimetre of conductor. As it has been inferred that the presence of the conductor is not essential for the production of an E.M.F., we may now vary this statement and say that a magnetic flux of density B moving with velocity u , produces an electromotive force of $Bu \times 10^{-8}$ volts per centimetre. The volt per centimetre is one of the units of electric field strength, hence the above argument can be expressed algebraically in the form

$$\Gamma = \frac{Bu}{c} = \frac{\mu H u}{c}$$

where c is a constant which has been introduced in order that the electric field strength Γ may be expressed in electrostatic units. It is the ratio of the electromagnetic unit of quantity to the equivalent electrostatic unit, or 3×10^{10} . The relative directions of B , u and Γ are shown in fig. 8a. The relation $\Gamma = \frac{Bu}{c}$ is sometimes called the second law of electro-dynamics.

Production of magnetic field by electric field in motion

24. Instead of commencing our reasoning from Faraday's law, it would be equally justifiable to commence by contemplation of an ordinary electric conduction current. This current is due to the movement of electric charges and its intensity is the rate at which the charges move. Now electric charges carry with them tubes of electric flux, one unit tube per unit charge. An electric current is, therefore, equivalent to a moving electric flux, and instead of the usual statement that an electric current produces a magnetic field, it may be asserted that an electric

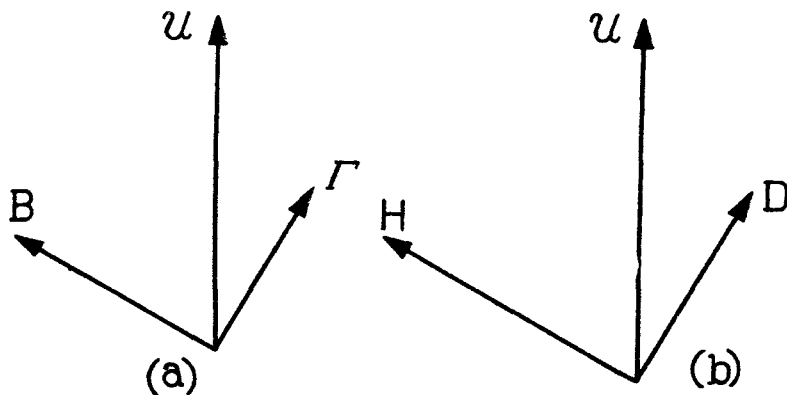


FIG. 8, CHAP. VII.—Laws of electro-dynamics.

flux in motion produces a magnetic field. We have previously also made use of the conception of magnetomotive force, the M.M.F. along a line of magnetic force being the work done in moving a unit pole along the line, that is to say, M.M.F. = force $\times$ distance, or M.M.F. = Hl . If this path encircles a conductor carrying a current of i E.M. units the M.M.F. is $4\pi i$ E.M. units. As was done earlier in the discussion, we may now substitute for the conduction current a displacement current in a small area A of dielectric material, and in this dielectric a small charge of Q units will carry Q tubes of flux, the average current during a time t being $\frac{Q}{t}$ electrostatic units. The area A may be considered rectangular and of dimensions l centimetres perpendicular to the

CHAPTER VII.—PARAS. 25-26

direction of current and x centimetres in the direction of the current, in which circumstances the current is $\frac{Q}{t} = \frac{DA}{t} = \frac{Dlx}{t}$. But $\frac{x}{t}$ is the average velocity u of the charge in the direction x hence $i = Dlu$ electrostatic units. Collecting our equations we may now write :—

$$\text{M.M.F.} = Hl = 4\pi i$$

$$Hl = 4\pi Dlu$$

$$H = 4\pi Du$$

except that the constant c must be inserted in order that H may be expressed as usual in electromagnetic units. Hence $H = \frac{4\pi Du}{c} = \frac{\kappa \Gamma u}{c}$ because $4\pi D = \kappa \Gamma$. The relative directions of D , H and u are shown in fig. 8b.. The relation $H = \frac{4\pi Du}{c}$ is called the first law of electrodynamics.

Self-supporting electromagnetic field

25. We now see that a moving electric field of a given strength will produce a magnetic field of definite strength, and that the converse is also true, a moving magnetic field producing a corresponding electric field. If we commence by postulating an electric field, which appears to be the more fundamental one since it is inseparable from the electrons of which matter is partly composed, we see that the motion of an electron gives rise to a magnetic field, and many examples of the manner in which this magnetic field is detected and utilised have been given in the earlier chapters. If the velocity of our fundamental field is sufficiently great, say u , the magnetic field produced by its motion will produce a new electric field of intensity equal to the original, and any change in the electrical or magnetic state of the medium at one point must be communicated to all points in the medium with the velocity u , at which the electric and magnetic fields become self-supporting. This velocity can be found from the equations developed above, for the magnetic field strength is then

$$H_1 = \frac{\kappa \Gamma_1 u_1}{c}$$

$$\text{but } \Gamma_1 = \frac{\mu H_1 u_1}{c}$$

$$\therefore H_1 = \frac{\kappa \mu u_1^2 H_1}{c^2},$$

and this equation can only be true if

$$\kappa \mu \frac{u_1^2}{c^2} = 1$$

$$\text{or } u_1 = \frac{c}{\sqrt{\kappa \mu}}.$$

Now in ether unobstructed by matter, $\mu = 1$ and $\kappa = 1$, and, therefore, $u_1 = c = 3 \times 10^{10}$ centimetres per second, which is identical with the velocity of light as determined by experiment. For this reason, among others, we conclude that light radiation is fundamentally an electromagnetic phenomenon.

The production of an electromagnetic wave

26. Consider a large rectangular loop of wire arranged in a vertical plane, say 100 yards in length and 10 yards high. If a source of unvarying E.M.F. is inserted in series with this loop,

a direct current will flow in it. The length of the loop has been made large in order that the mutual effects of the vertical portions of the wire may be neglected, and it is proposed to consider the current flowing in only one of the vertical portions, namely, that in which the electron flow has an upward direction. Let us further confine our observations to the phenomena associated with a single electron, which is an electric charge of e units, and may be considered to have a constant velocity of about one centimetre per second, which will be denoted by b . The electron is the focus of an electric field consisting of lines of electric force which converge upon it from all directions and extend to an infinite distance, this field moving with the electron and having the same velocity through space. At any point in this field distant x centimetres from the conductor the electric flux density will be $\frac{e}{4\pi x^2}$. (Chapter I.) In order to simplify matters still further, the behaviour of a single line of force will be observed. The relevant portion of the circuit is then as illustrated in fig. 9 where M represents the position of the electron at a certain instant, and the single line of force under observation is represented by Mm.

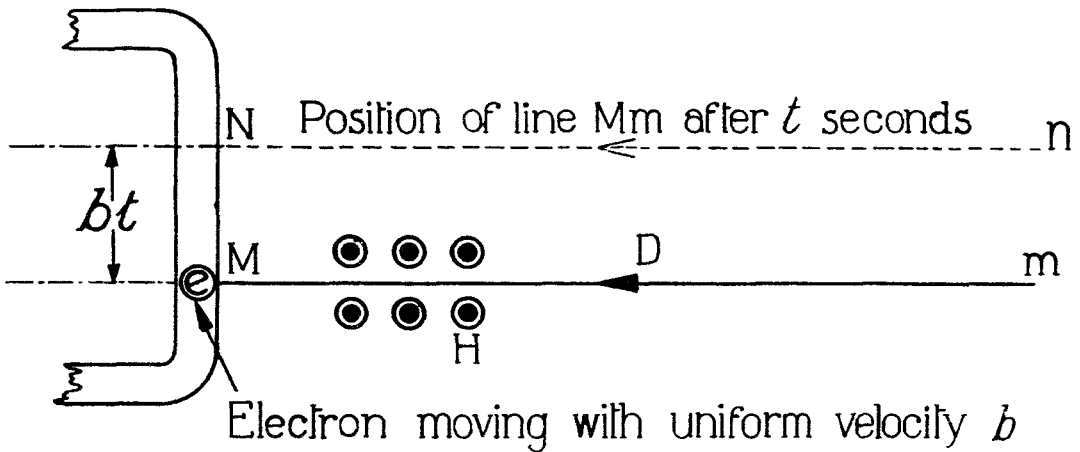


FIG. 9, CHAP. VII.—Field due to electron moving with constant velocity.

27 Owing to the motion of the line of force, a magnetic field will be produced, as described in the preceding section. This magnetic field will be directed perpendicularly to the direction of motion of the line, and also perpendicularly to the direction of the electric field, hence by the first law of electro-dynamics the magnetic field strength H at all points in the line Mm is in the direction "out of the paper." This may be verified by the corkscrew rule, observing, however, that the conventional "direction of current" used in this rule is opposite to that in which the electron movement is taking place. In accordance with the foregoing analysis the strength H of this magnetic field is $4\pi D \frac{b}{c}$ E.M. units; as $4\pi D = \frac{e}{x^2}$, $H = \frac{e}{x^2} \cdot \frac{b}{c}$, and it is demonstrated that the magnetic field strength caused by the moving line of force varies inversely as the square of the distance from the conductor, but is of constant magnitude at any point if the velocity b is constant. This moving magnetic field in its turn generates an electric field, and its intensity by the foregoing analysis is $\mu H \frac{b}{c}$ or $\frac{\mu e}{x^2} \cdot \frac{b^2}{c^2}$. Its magnitude is extremely small compared with the original electric intensity and may be disregarded.

28. The motion of an electron or any number of electrons with constant velocity in the conductor thus gives rise to a magnetic field in the surrounding medium, which forms concentric

CHAPTER VII.—PARA. 29

circles round the conductor, and occupies the same space as the radial electric field. These fields are called the induction fields and their intensity varies inversely as the square of the distance from the conductor. In accordance with the conclusions reached in paragraph 25, these fields would be self-sustaining if the velocity of the electron in the first place was sufficiently great,

that is, if $b = \frac{c}{\sqrt{\kappa\mu}}$ or as the values of κ and μ for air are to all intents and purposes unity, if

$b = c$. Thus it appears that if an electron were moving in the conductor with velocity c , and its motion were suddenly arrested, the fields would continue in motion in the upward direction, forming a single pulse which would travel into space. This speculation is an idle one, however, firstly, because the above reasoning is only strictly applicable when b is very small compared to c , and secondly, because in order to bring its velocity to zero the electron must receive acceleration. It will now be shown that the effect of such an acceleration will be to produce in the medium a self-sustaining electromagnetic field, or what is generally called an electromagnetic wave, which is propagated *radially outward* from the conductor.

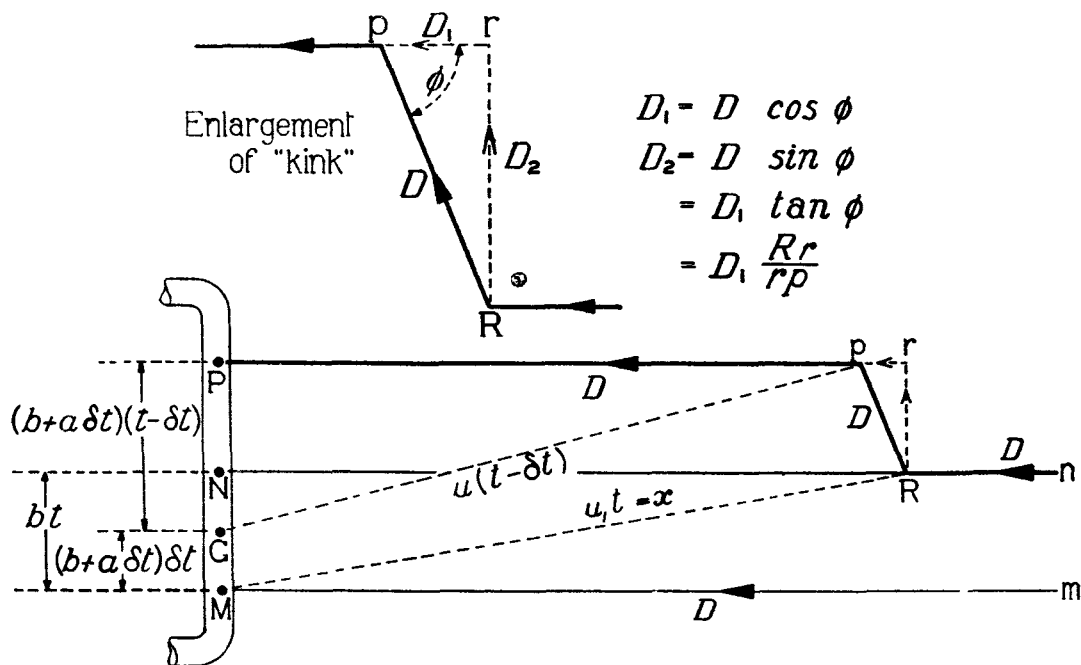


FIG. 10, CHAP. VII.—Field due to electron moving with accelerated velocity.

29. Let the conditions already laid down be repeated, the electron travelling upward, first with uniform velocity b from M , so that the electric line of force under observation is Mm . After t seconds the electron reaches the position N where $MN = bt$, and the line of force will move with it, reaching the position Nn . This repetition of the previous instance has been performed merely to serve as a frame of reference on the diagram, fig. 10. Now suppose that on arrival at the point M the electron receives an acceleration a which lasts for a short interval of time δt after which the electron continues to travel with its new velocity, $b + a\delta t$. In the time t , the electron will arrive at the point P instead of N , but as disturbances in the medium are propagated

with the finite velocity $u_1 = \frac{c}{\sqrt{\kappa\mu}}$, the occurrence of the acceleration is only effective upon that portion of the line of force which is within a radius $u_1 t$ centimetres from the point M.

30. (i) The portion Rn of the original line of force, which is situated outside the radius $u_1 t$, is still a portion of the line of force under consideration. Within the radius $u_1 (t - \delta t)$ from the point G at which the acceleration ceased, however, the line of force assumes the position Pp, and the whole line of force instead of being Nn, has been changed by the acceleration into the line Pp Rn which is called a "kinked line" for the sake of brevity. The electric flux density D at the point p can be resolved into two perpendicular components, the first being in the radial direction r p and the second in the direction of motion of the line of force. Denoting the angle rpR by ϕ , and remembering that the radial electric flux density D_1 at the point p is $\frac{e}{4\pi x^2}$,

$$\begin{aligned} D_1 &= D \cos \phi \\ D_2 &= D \sin \phi \\ &= D_1 \frac{\sin \phi}{\cos \phi} \\ &= D_1 \tan \phi \\ &= D_1 \frac{Rr}{rp} \end{aligned}$$

(ii) To obtain values for Rr and rp it must be observed that by hypothesis the velocity of the electron, b , is very much smaller indeed than the velocity u_1 , so that in fig. 10, MR may be regarded as equal to NR and Gp equal to Pp. Then rp, which is actually equal to NR - Pp, may be said to be equal to MR - Gp, or since MR = $u_1 t$ and Gp = $u_1 (t - \delta t)$, to $u_1 \cdot \delta t$. Similarly Rr is equal to NP and NP = GP + GM - MN.

$$\begin{aligned} \text{Now } GP &= (b + a \cdot \delta t)(t - \delta t) \\ &= bt - b \cdot \delta t - a \cdot \delta t^2 + at \cdot \delta t \\ GM &= (b + a \cdot \delta t) \delta t \\ &= b \cdot \delta t + a \cdot \delta t^2 \\ MN &= bt \\ \therefore NP &= bt - b \cdot \delta t - a \cdot \delta t^2 + at \cdot \delta t + b \cdot \delta t + a \cdot \delta t^2 - bt \\ &= at \cdot \delta t, \\ \text{and } D_2 &= D_1 \frac{a t \cdot \delta t}{u_1 \cdot \delta t} \\ &= D_1 \frac{a \cdot t}{u_1} \\ &= D_1 \frac{at u_1}{u_1^2} \\ \text{but } u_1 t &= x \\ \therefore D_2 &= D_1 \frac{a x}{u_1^2} \end{aligned}$$

(iii) The velocity of the component of flux D_2 is in the direction p r, i.e. radially outward, and is equal to u_1 centimetres per second, while the velocity of the component D_1 is equal to that of the electron, namely b centimetres per second, and is in the upward direction. The total

CHAPTER VII.—PARA. 30

electric flux density is the vector sum of these, and the corresponding electric field strengths can be calculated by the relation $\Gamma = \frac{4\pi D}{\kappa}$ giving the following results, in dynes per unit charge :—

$$\begin{aligned}\Gamma_1 &= \frac{e}{\kappa x^2} \\ \Gamma_2 &= \frac{a \kappa}{u_1^2} \Gamma_1 \\ &= \frac{a \kappa e}{\kappa x^2 u_1^2} \\ &= \frac{a e}{\kappa x u_1^2}\end{aligned}$$

by noting that $\frac{1}{u_1^2} = \frac{\kappa \mu}{c^2}$ this becomes

$$\Gamma_2 = \frac{a e u}{\kappa c^2}$$

Γ_1 , which varies inversely as the square of the distance from the conductor, denotes the strength of the induction field, while Γ_2 , which varies inversely as the distance, is the strength of the radiation field. Each of these has an associated magnetic field, for the total magnetic field may also be resolved into two components, one caused by the movement of the electron with constant upward velocity b , the other caused by the motion of the electric radiation field with radial velocity u_1 . These have the following magnitudes in dynes per unit magnetic pole :—

$$\begin{aligned}H_1 &= \frac{\kappa \Gamma_1 b}{c} \\ &= \frac{e b}{x^2 c}\end{aligned}$$

which is the induction field strength, while the radiation field has the strength

$$H_2 = \frac{e a}{x} \frac{\sqrt{\kappa \mu}}{c^2}$$

In free space or any other medium of which the dielectric constant κ and magnetic permeability μ are unity.

$$\Gamma_2 = \frac{a e}{x c^2}$$

$$H_2 = \frac{a e}{x c^2}$$

or $\Gamma_2 = H_2$ numerically, although their units are different.

The above results are summarised in the following table.

Motion of charge	Radial electric field Γ_1	Electric field Γ_2 perpendicular to Γ_1 and x	Magnetic field H perpendicular to Γ_1 and Γ_2
Nil	$\frac{e}{\kappa x^2}$	Nil	Nil
Constant velocity b	$\frac{e}{\kappa x^2}$	Nil	$H_1 = \frac{e b}{x^2 c}$
Velocity b , Acceleration a	$\frac{e}{\kappa x^2}$	$\frac{e a \mu}{x c^2}$	$\frac{e b}{x^2 c} + \frac{e a \sqrt{\kappa \mu}}{x c^2}$ $= H_1 + H_2$

31. Now if instead of one electron undergoing a single acceleration, we consider the effect of the whole electron stream in the conductor, it is obvious that the intensity of the radiated field will be directly proportional to the current, and the effect of a single acceleration, e.g. that caused by switching the current on or off, will be to radiate a single pulse which can be detected by suitable apparatus at considerable distances. Such a pulse is of little value for ordinary signalling, however, and it is preferable to initiate a train of such pulses, similar to the waves in water to which reference has already been made. In order to produce electro-magnetic waves of this nature the electrons in the conductor must undergo sinusoidal acceleration and the resulting wave will then consist of sinusoidal magnetic and electric fields, which are perpendicular to each other in space. As already shown, the radiation electric field is parallel to the wire carrying the current to which the wave owes its existence, and the radiation magnetic field is perpendicular to the electric field, while both these fields are perpendicular to the direction in which the wave is travelling. This orientation of the fields, with respect to the conductor carrying the current by which the wave is caused, may be referred to as the natural polarisation of the wave. In the immediate vicinity of the conductor, the induction field is stronger than the radiation field; the former varies inversely as the square of the distance, and the latter inversely as the distance from the conductor, the total field being the vector sum of the two, so that very near to the conductor the total magnetic and electric fields are very nearly 90° out of phase with respect to time. As we go further and further from the conductor, the induction field strength falls off rapidly, and at a distance $\frac{\lambda}{2\pi}$ the induction field has fallen to equality with the radiation field.

At still greater distances the induction fields become negligible and the magnetic and electric fields are in phase with each other.

Polarisation of wave

32. Suppose that an electromagnetic wave is originated as a result of the sinusoidal acceleration of electrons in a vertical wire, situated near the surface of the earth. The wave will be emitted with natural polarisation, and will travel over the surface of the earth with its magnetic field in the horizontal and its electric field in the vertical plane. From the point of view of reception of such waves, we are not particularly interested in the natural polarisation, but in the manner in which the wave is incident upon the receiving aerial. It is, therefore, usual to state the polarisation with reference to the earth's surface, and a wave which reaches the receiver in such a manner that the magnetic field is in the horizontal plane, and the electric field in the vertical plane is said to be normally polarised. If the orientation of the fields differs from this in any way whatever, the wave is said to be abnormally polarised, and the angle of polarisation is defined as the angle which the electric field makes with the vertical plane.

33. The practical production of an electro-magnetic wave suitable for radio-telegraphic purposes is contingent upon the sinusoidal acceleration of electrons and owing to the properties of simple harmonic motion (paragraph 10) this indicates the production of a sinusoidally varying current. We have seen that in simple harmonic motion the acceleration is proportional to the square of the frequency, hence in order to obtain appreciable radiation from the circuit, the frequency must be high. It is possible to provide such an alternating current, having a frequency of the order of 30,000 cycles per second, by means of an alternator constructed on the principles discussed in Chapter IV, but such machines have very little to commend them on practical grounds. The desired result can be obtained with much greater convenience by the use of an oscillatory circuit as described in the earlier paragraphs of this chapter.

Radiation from closed loop

34. It has already been stated that the type of oscillatory circuit which is best adapted for the radiation of energy in the form of electromagnetic waves is an open oscillator, which may be described as a circuit possessing inductance and capacitance and having an inherent resistance which is much lower than the critical value, the geometric dimensions of the circuit being of the same order as the wavelength of the oscillation; the last stipulation ensures that the inductance

CHAPTER VII.—PARA. 35

and capacitance of the circuit are distributed in space and not localised as in the closed oscillator. In order to illustrate the effect of the geometry of the circuit upon the amount of energy radiated let us again contemplate the vertical rectangular loop of a preceding paragraph, but now consider two electrons, one being situated in each vertical side. In fig. 11 the point P is supposed to be situated at a considerable distance from the loop on the side opposite to the reader, who is supposed to observe the electric fields at P by looking through the loop. When a current is established in the circuit, the electron e_1 is moving upwards and the electron e_2 downwards, and if both receive equal positive acceleration, or increase of velocity, at the same instant, similar kinks will travel outward from each electron. Considering only the lines of force passing through P, it is seen that the kink emanating from the electron e_1 will reach this point at exactly the same moment as the kink originated by the acceleration of e_2 , but as the latter acceleration is in the downward direction while that of e_1 is upward the outward-travelling flux-density (both magnetic and electric) of the two kinks are in antiphase and their combined effect is to annul each other, hence the effective travelling flux at the point P is zero, no matter how near to the loop P may be, provided it is equidistant from e_1 and e_2 . At any point in the direction X, however, the kink due to the acceleration of e_1 is not received until a time $t = \frac{x}{c}$ after the effect of the acceleration of e_2 , because the kink caused by e_1 has to travel a greater distance, i.e. the width of the loop,

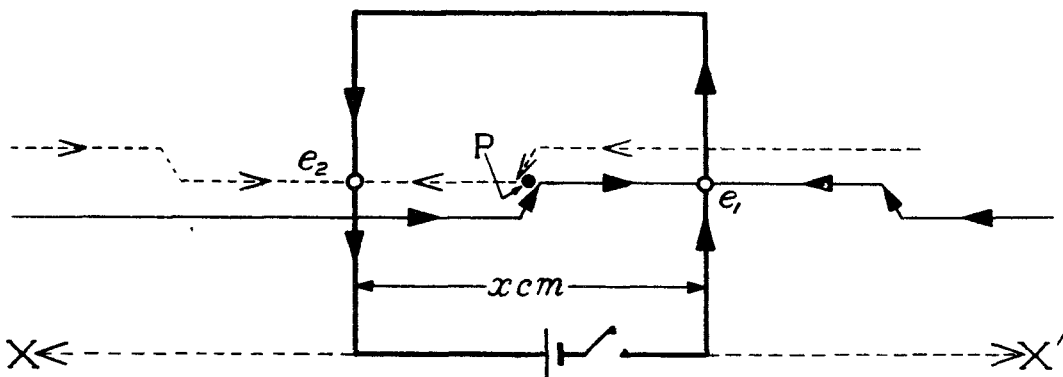


FIG. 11, CHAP. VII.—Radiation from closed loop.

which is x cms. (see fig. 11.). Similar considerations apply to the direction X^1 , and it can be seen that a loop of this nature, i.e. in which the sides are an appreciable distance apart, will have directional radiating properties, which will receive further consideration in Chapter XVI. For the present, suppose that the width of the loop x is reduced. The difference in the time of arrival of disturbances due to e_1 and e_2 at any point in XX^1 , will be less than before, and eventually as x approaches zero the amount of energy radiated along XX^1 becomes negligible. Thus, the effect of bringing the two sides into proximity so that the loop becomes non-inductive, is to render it non-radiative also.

Radiation of damped waves from dipole

35. Instead of a closed loop, let only a single vertical wire be used, which we will first suppose to be located well above the ground, in order that the capacitance of the wire with respect to earth shall have an inappreciable effect upon the electrical conditions. It will also be assumed that it is possible to locate a battery and switch in a position near the mid-point of the wire. Although a circuital conduction current cannot be established, it is still possible to cause an acceleration of the electrons in the vertical wire, because the latter possesses a certain amount of capacitance, the length connected to the positive terminal and that connected to the negative forming a kind of condenser. On closing the switch S (fig. 12) a momentary charging current will

flow in accordance with the principles explained in Chapter I, and consequently all the electrons in the conductor are momentarily displaced, i.e. they receive acceleration. This charging of the capacitance will be followed by a discharge which will consist of a damped train of oscillations as already explained. The vertical wire has, in fact, become a rudimentary form of transmitting aerial and will radiate equally well in all directions in the horizontal plane. If L is the effective inductance, C the effective capacitance and R the effective resistance of the wire the frequency of the oscillations will be given by the formula $f_n = \frac{1}{2\pi} \sqrt{\frac{1}{LC} - \frac{R^2}{4L^2}}$, and the log. dec. of the oscillations by $= \pi R \sqrt{\frac{C}{L}}$. The effective inductance and capacitance obviously depend upon the

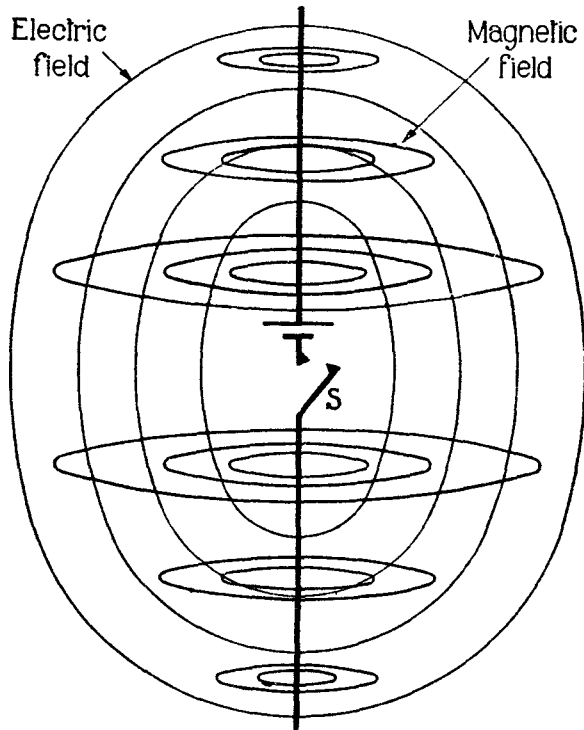


FIG. 12, CHAP. VII.—Electric and magnetic fields of dipole.

length of the wire, and it is found that under the conditions laid down, the wavelength of the emitted radiation is twice the length of the wire, which may be expressed algebraically as $\lambda = 2l$, where λ is the wavelength and l the total length of the wire. An aerial of this form is, therefore, called a half-wave aerial or a dipole. The latter term may be interpreted as signifying that the lengths of wire connected to each terminal are merely devices for increasing the inductance and capacitance of the two "poles" of the battery.

36. Single trains of waves produced by a battery in the manner described are, of course, useless for radio telegraphic communication, unless the wave trains are of comparatively long duration and are caused to succeed each other at very short intervals, while in order to initiate appreciable radiation, a much larger quantity of energy must be stored in the aerial capacitance than is practicable by means of a battery as envisaged above. As our present object is to discuss

CHAPTER VII.—PARA. 37

the application of electromagnetic radiation to communication, it is of interest to calculate the duration of a wave train such as would be produced by the simple apparatus described above. For this purpose we shall require to know the decrement of the circuit. This has been defined as

the ratio $\frac{I_1}{I_2} = \frac{I_2}{I_3} = \frac{I_3}{I_4}$, where I_1, I_2 , etc., are the amplitudes of successive peaks of oscillatory current of the same sign. The wave train may be considered to carry negligible energy when the amplitude has fallen to .01 of the amplitude of the first peak. Now $\frac{I_1}{I_2} \times \frac{I_2}{I_3} \times \frac{I_3}{I_4} = \frac{I_1}{I_4} = N^3$

where N is the decrement, and must not be confused with the log. dec. Similarly $\frac{I_1}{I_2} \times \frac{I_2}{I_3} \times \frac{I_3}{I_4} \times \frac{I_4}{I_5} = \frac{I_1}{I_5} = N^4$ and by extending this process to any number of successive amplitudes it is apparent

that, if I_n represents the amplitude of the n th peak $\frac{I_1}{I_n} = N^{n-1}$, hence if $I_n = .01 I_1$, $N^{n-1} = 100$. This equation can be solved by taking the logarithms of both sides. If common logs. are used, $\log_{10} 100 = 2$ while $\log_{10} N^{n-1} = (n-1) \log_{10} N$. We do not yet know the value of $\log_{10} N$, but it can easily be found, if we know the naperian log. of N , i.e. δ ; the common log. is $\frac{1}{2.3026}$ or .4343 of this, that is .4343 δ . Hence

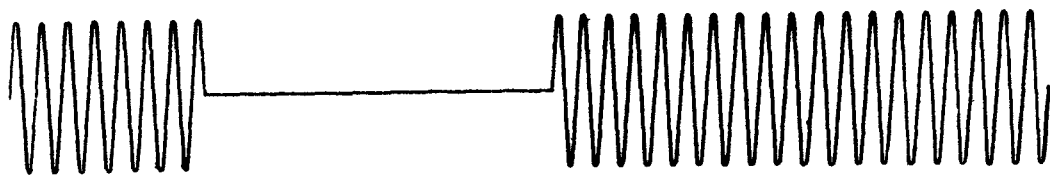
$$\begin{aligned} 2 &= (n-1) \log_{10} N \\ \frac{2}{\log_{10} N} &= n-1 \\ \therefore n &= 1 + \frac{2}{.4343\delta} \\ &= 1 + \frac{4.605}{\delta} \end{aligned}$$

Taking typical values for the constants of a vertical aerial wire 10 metres or so in length, $L = 20 \mu H$, $C = .00005 \mu F$, $R = 100$ ohms

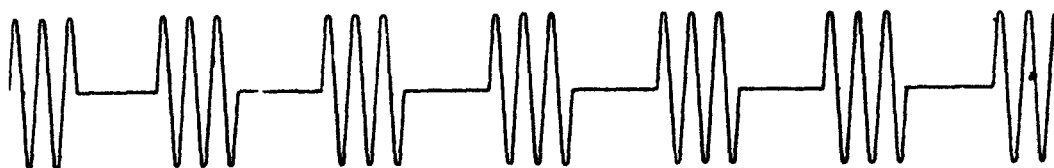
$$\begin{aligned} \delta &= \pi R \sqrt{\frac{C}{L}} = 314 \sqrt{\frac{.00005}{20}} = .496 \\ n &= 1 + \frac{4.605}{.496} \\ &= 1 + 9.3 \\ &= 10.3 \end{aligned}$$

The amplitude of the 11th wave will be less than .01 of the initial amplitude. The number of waves, multiplied by the period of each wave will give the duration of the wave train. The period is $2\pi\sqrt{LC}$ (approximately) or 1.985×10^{-7} seconds, and, therefore, the whole wave train will last only about 2×10^{-6} seconds.

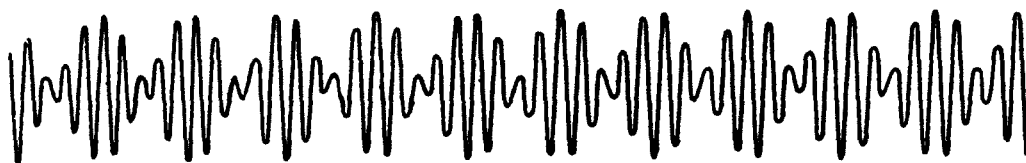
37. (i) If it is desired to signal with the morse code, using aural reception (i.e. by employing the telephone head set) the duration of a "short" must be not less than about $\frac{1}{25}$ th second, and a "long" about $\frac{1}{5}$ th second. Wave trains of such brief duration as 2×10^{-6} second can only be utilised if arrangements are made for the charging and discharging processes to take place with extreme rapidity, so as to produce, say 2,000 complete wave trains per second. Another reason why damped wave trains are unsuitable for communication purposes is the difficulty of radiating a large quantity of energy. The energy stored in the capacitance of the vertical aerial is $\frac{1}{2} C \mathcal{V}^2$ joules, and if arrangements were made to charge and discharge the aerial n times per second the total power expended would be $\frac{1}{2} C n \mathcal{V}^2$ joules per second or watts.



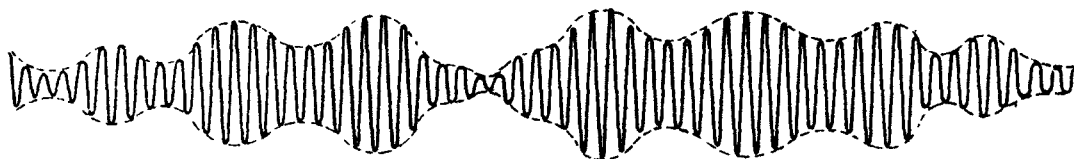
Type A1 wave



I.C.W. wave



M.C.W. or T.T.
Type A2 waves



Modulated wave

Type A3.modulation frequency within audible limit

Type A4.modulation frequency above audible limit

TYPES OF RADIO WAVES

FIG. 13
CHAP. VII

Example 3.

An aerial of capacitance $0.00005 \mu F$ is charged to a P.D. of 2,000 volts and then allowed to discharge, the process being repeated 2,000 times per second. What is the power supplied to the oscillator?

$$\begin{aligned}
 P &= \frac{1}{2} C n V^2 \\
 &= \frac{1}{2} \times 5 \times 10^{-11} \times 2,000 \times 2,000^2 \\
 &= \frac{1}{2} \times 5 \times 10^{-11} \times 2 \times 2 \times 2 \times 10^9 \\
 &= 20 \times 10^{-2} \\
 &= \cdot 2 \text{ watt.}
 \end{aligned}$$

(ii) Of the power supplied, less than 50 per cent. may be converted into radiation (see paragraph 51). The amount of energy stored in the aerial can only be increased by increasing its capacitance, raising the voltage, or both. The capacitance must depend to some extent upon the frequency chosen for communication while the increase of voltage leads to other difficulties. Nevertheless the original "spark" or damped wave telegraphy was achieved in this manner, which however has been rendered quite obsolete by the development of continuous or undamped wave telegraphy, while the technique of radio-telephonic transmission depends essentially upon the production of an undamped oscillation.

38. In order to produce such an oscillation it is necessary to introduce into the oscillatory circuit during some portion of every cycle, an amount of energy equal to that which has been expended during the preceding cycle. If the amplitude of the first peak of oscillatory current is $\mathcal{A}$ amperes and the total losses are represented by an effective resistance of R ohms, the energy expended during the first half-cycle is $\frac{\mathcal{A}^2 R}{4 f_n}$ joules, and if an amount of energy equal to this is supplied to the circuit during the half-cycle immediately following the attainment of the first peak value, the next peak of current, which will be of negative sign, will reach the same amplitude as the first peak. If this addition of energy is performed at some period during every succeeding cycle an undamped oscillation will be produced, while if the energy added is in excess of the amount dissipated in various losses during the cycle, the amplitude of oscillation of each succeeding cycle will increase until some peak value is reached at which the energy supplied per cycle is just sufficient to make good the damping losses. A close analogy to this is the oscillation of a swing, which is maintained (and increased in amplitude if desired) by successive pushes at the moment when the swing has just passed the peak of its oscillation.

Wave form of electromagnetic waves

39. (i) Electromagnetic waves have been divided into several classes according to the waveform of the disturbance, which in turn depends upon the manner in which the oscillatory circuit is supplied with energy. There are two main classes:—

Type A waves, which are of constant frequency and amplitude and are therefore called "continuous waves".

Type B waves, which are the damped waves emitted by an oscillator of which the capacitance is charged some 100 to 2,000 times per second and then allowed to discharge at its natural frequency, as described in paragraph 37. Type B waves are no longer used for communication purposes in the R.A.F.

Type A waves are only of use for such purposes as radio beacons, which enable an aircraft carrying the necessary apparatus to obtain its bearing from the transmitter. For signalling purposes, it is necessary to interrupt the emission in accordance with the morse code, or to vary its amplitude as the magnitude of the direct current in a simple telephone circuit is varied by the carbon microphone. In the former case, the waves are said to be key controlled, and in the latter to be modulated. The complete division of Type A waves is therefore as follows:—

Type A1 waves.—Continuous waves, unmodulated, key controlled. These are continuous waves of which the amplitude or frequency (or both) are varied by the operation of keying, for the purpose of telegraphic communication.

CHAPTER VII.—PARAS. 40-41

Type A2 waves.—Interrupted continuous waves (I.C.W.). Continuous waves in which a variation of amplitude is made in a periodic manner at audio frequency, and key controlled for the purpose of telegraphic communication (W/T). When the variation of amplitude is approximately sinusoidal, the emission is referred to as "Tonic train" (T.T.).

Type A3 waves.—Sound modulated waves, continuous waves in which a variation of amplitude or frequency is made in accordance with the characteristic vibrations of speech and music (R/T).

Type A4 waves.—Modulated waves, Continuous waves in which a variation of amplitude or frequency is made; the modulation frequencies may be very much higher than those required for the transmission of speech and music. This form of wave is utilised for television transmission.

(ii) Fig. 13 shows the various forms of electro-magnetic wave diagrammatically. In the case of Type A1 waves a morse "short" and "long" are shown, but it must be realised that many more waves are required to constitute each morse element than are actually drawn. If the "short" lasts $\frac{1}{20}$ th second and the frequency of the oscillation is 10^6 cycles per second, the "short" would consist of 50,000 complete waves. In the representation of Type A2 waves, only a single "short" is shown. If the interruption takes place 200 times per second, the short will consist of ten groups of waves, and as a "long" has three times the duration of a "short" the "long" would consist of thirty groups. The representation of the Type A3 wave shows how the amplitude of the wave varies in accordance with the vibrations of speech or music. The dotted line which is touched by the peak of each successive wave must be regarded as merely a "construction line". It is the wave form of the speech or music which is being transmitted, and is often referred to as the "modulation envelope" of the wave. The significance of this envelope will be dealt with in the chapter devoted to radio-telephony.

PRACTICAL RADIATING CIRCUITS OR AERIALS

40. In modern practice the introduction of energy into the oscillatory aerial circuit is accomplished by means of a thermionic valve, which is so connected as to draw a supply of energy from a source of constant E.M.F. and inject this energy into the oscillatory circuit during every cycle. The oscillatory current in the aerial circuit is thus maintained at a constant amplitude, and the valve, with its supply batteries, may be regarded simply as an alternator, the frequency of which accommodates itself to the natural frequency of the aerial circuit. The phenomena associated with the thermionic valve itself, and the manner in which the valve functions when acting as an energy converter, are discussed in Chapters VIII and IX. In the succeeding paragraphs, the source of energy in the oscillatory circuit will be considered to be an alternator of zero internal impedance, the frequency of which is coincident with the natural frequency of the circuit.

Stationary waves on conductors

41. Suppose, therefore, that a long wire is suspended vertically some distance above the ground as before, but with a source of alternating E.M.F. of any frequency whatever connected in place of the battery. The wire may be considered to have its inductance and capacitance equally distributed throughout its length, and its resistance will be neglected. The application of an alternating E.M.F. then causes an acceleration of the electrons situated in those portions of the wire which are immediately adjacent to the terminals of the alternator, the resulting movement of the electrons will cause repulsion of those in the vicinity, and the latter also receive acceleration. In a loss-free conductor the acceleration of electrons is, in fact, communicated from one to the other in the length of the conductor with a velocity which is found to be 3×10^{10} centimetres per second, precisely that with which electromagnetic disturbances are propagated in free space, although if the effect of the resistance of the wire is taken into consideration it may be somewhat less. The reader may appreciate a reminder that the velocity with which the

acceleration is communicated from electron to electron (3×10^{10} centimetres per second) is not the velocity with which the electrons actually move along the wire, the latter being only of the order of one centimetre per second, and if the two statements appear to conflict, the position may be cleared up by an analogy. When an engine starts to move a long train of wagons from a standstill the velocity with which the acceleration is communicated may be roughly estimated by measuring the time interval between the successive clanking sounds which accompany the acceleration of each wagon. If this is half a second, and each wagon is 22 ft. long, the acceleration is communicated from wagon to wagon with a velocity of $22 \div \frac{1}{2}$ or 44 ft. per second. The whole of the train is not in motion until the last wagon has received acceleration, and it is absurd to suppose that because the acceleration is conveyed with a velocity of 44 ft. per second the train is moving at 30 miles per hour!

42. We may, therefore, conclude that there is nothing incompatible in the statements that the average velocity of the electrons is about one centimetre per second, and that the occurrence of acceleration is communicated from electron to electron with a velocity approaching that of light. The fact that a variation of E.M.F. is occurring, then, is the cause of a progressive disturbance of electrons in the wire, and by our definition of wave motion we may say that an electric wave exists in the wire. This wave starts at the terminals of the alternator and travels outwards along each wire, but on reaching the ends of the latter, is reflected, and the wave travels back along the wire to the alternator, impressing an acceleration in the reverse direction upon the electrons. The acceleration of some electrons in the wire may, therefore, be the sum of two equal and opposite accelerations, or zero. Since the current is proportional to the velocity of the electrons, a current wave is also set up in the conductor. Confining our attention to the latter and assuming that no reflection occurs at the generator itself, the waves reflected at the ends of the wire travel back towards the opposite ends and a very complex electrical state may exist, but however complex it may be, the current at the ends of the wire remote from the generator is always zero, for no electrons can travel past these points. If the frequency of the supply is such that the length of wire attached to each terminal of the generator is an exact multiple of one-quarter of a wavelength (the wavelength being related to the frequency by the relation already given, viz., $\lambda = \frac{c}{f}$) the direct and reflected waves combine in such a manner that no travelling wave exists, but instead, what is called a standing or stationary wave is set up in the wire.

43. Without attempting academic accuracy, we can see that such a wave will be produced by reflection if the original current is assumed to be $\mathcal{I} \sin \omega (t - kx)$, that is to say, a current of the form $\mathcal{I} \sin \omega t$ which varies in phase from point to point in space, the latter being by definition a travelling wave. On reflection without loss of amplitude, the wave becomes $\mathcal{I} \sin \omega (t + kx)$, the positive sign signifying that the reflected wave is travelling in a direction opposite to the first, and the current at any point x is

$$\mathcal{I} \sin (\omega t - k_1 x) + \mathcal{I} \sin (\omega t + k_1 x),$$

the new constant k_1 being simply ω times the constant k previously used. The sum of these sinusoidal quantities, according to a formula developed in Chapter V, is

$$\begin{aligned} & \mathcal{I} (\sin \omega t \cos k_1 x + \cos \omega t \sin k_1 x) \\ & + \mathcal{I} (\sin \omega t \cos k_1 x - \cos \omega t \sin k_1 x) \\ & 2 \mathcal{I} \sin \omega t \cos k_1 x, \\ & \text{or } (2 \mathcal{I} \cos k_1 x) \sin \omega t. \end{aligned}$$

This indicates that the amplitude of the current at any point in the wire is proportional to the current at the middle of the circuit and also varies with the distance from the mid-point. There is, however, no part of the expression which signifies a change of phase from point to point in the wire, and we conclude that the direct and reflected wave have combined to form the stationary wave. In the kind of circuit under discussion, that is one which has inductance and capacitance

CHAPTER VII.—PARA. 44

distributed along the length of the circuit in a more or less uniform manner, resonance is said to exist when the applied E.M.F. causes stationary waves to be set up in the wire. The wavelength corresponding to any resonant frequency depends upon the length of the wire, the lowest resonant frequency being that which makes the aerial a half-wave aerial, having a current maximum—which is termed a current loop—at the centre, while the current at either end is zero. Points at which the current is zero are termed current nodes, and so, in a half-wave aerial, we have a current loop at the centre and current nodes at either end.

44. From another point of view, the variation of current in different parts of the aerial may be attributed to the existence of distributed inductance and capacitance along the length of the conductor, as shown diagrammatically in fig. 14a. It must be appreciated that instead of the finite number of elements of inductance and capacitance which have been drawn, an infinite number really exist. Suppose the current flowing through the generator to be i amperes, and its direction upwards, at a given instant. A portion of this current, i , will leave the conductor at the point P in the form of a displacement current through the element of capacitance C_1 , and

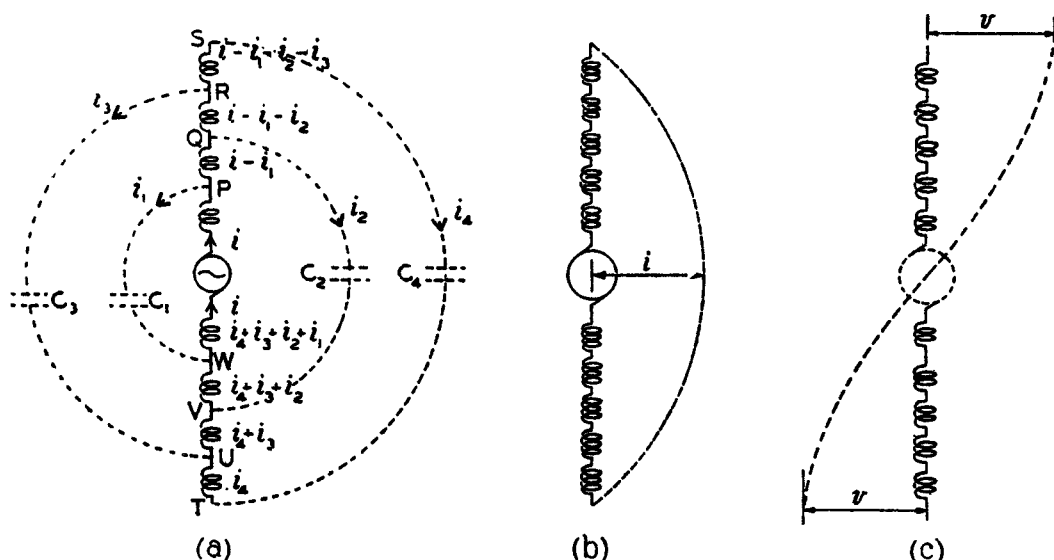


FIG. 14, CHAP. VII.—Current and voltage distribution of dipole.

the current flowing in the next element of inductance will be $i - i_1$. On reaching the point Q, a further displacement current i_2 will flow through the element of capacitance C_2 , and the current in the wire itself will be $i - i_1 - i_2$. This reduction of conduction current continues at all points in the wire, and the current at the end of the wire is zero as already stated. Now consider the length of wire attached to the other terminal of the generator. At the instant under consideration, the current in this wire is also upward and is zero at the extreme end, but at the point T the displacement current i_4 enters the wire and flows toward the generator, being joined at U by i_3 , at V by i_2 and so on. At the lower generator terminal therefore the current flowing upward is $i_1 + i_2 + i_3 + i_4 = i$ which is equal to that leaving by the upper terminal, and the amplitude of current in the aerial varies from a maximum at the centre to zero at the ends, fig. 14b. The P.D. with regard to the midpoint also varies from point to point in the wire, increasing as the distance from the midpoint increases, and being therefore greatest at the ends of the conductor. Thus all points which are loops of current are nodes of potential, while the positions of current nodes coincide with those of loops of potential. If then the conductor

acts as a half-wave aerial, it possesses a potential node at its centre and potential loops at the ends (fig. 14c). Diagrams showing the amplitude of current and P.D. at all points in the aerial are termed current and voltage distribution diagrams respectively.

Quarter-wave aerial—the counterpoise

45. The form of aerial hitherto considered, which is energised at its midpoint, is frequently found to be impracticable, owing to its excessive length when low-frequency radiation is desired. By a simple expedient the length of aerial required to radiate a given frequency can be reduced by one half. This expedient is the replacement of one half of the dipole by a conductor of as large an area as possible, and the capacitance of the aerial is then that existing between this area of metal and the remaining half of the wire (fig. 15a). This is a form eminently suitable for use in

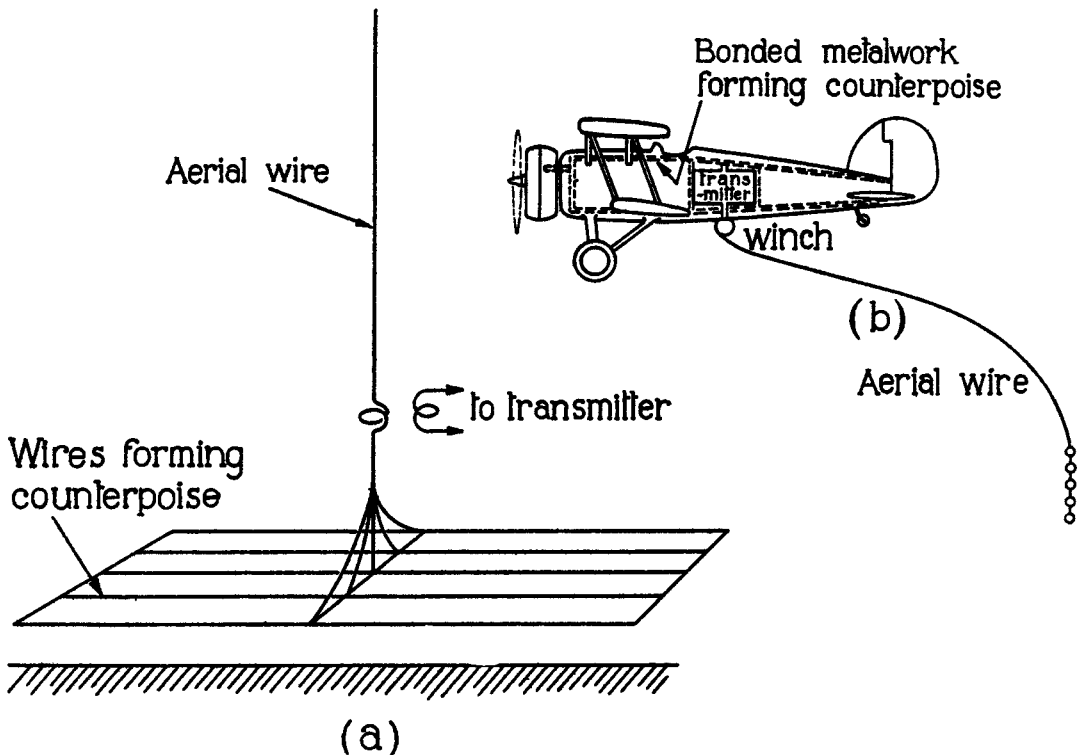


FIG. 15, CHAP. VII.—Counterpoise in ground station and aeroplane.

aircraft, for the actual aerial wire can be suspended from the aircraft and the "opposite plate" of the aerial capacitance is constituted by the whole of the metal work of the aircraft, which is held in electrical continuity by the use of bonding strips (fig. 15b). When operating at its natural frequency, one quarter of a stationary wave is set up in the aerial, a potential loop and current node existing at the lower end and a potential node and current loop at the point where it is connected to the bonding. The radio transmitter itself is connected between the latter point and the aerial wire. When used in the manner described, any conducting area is termed a counterpoise.

Earthed vertical aerial

46. When the radio transmitter is intended for use at a fixed location on the ground, an apparently obvious development is to make use of the earth itself as a counterpoise, and it now

CHAPTER VII.—PARA. 47

seems hardly credible that ten years elapsed between the original production of electromagnetic waves by means of a dipole, which was achieved by Hertz in 1884, and the introduction of the earthed vertical aerial which was due to Marconi. The latter is the simplest form of ground station aerial, and the whole aerial system then consists of a vertical wire one quarter of a wavelength in height, earthed at its lower and insulated at its upper ends. The high frequency power supply device—which is usually a triode oscillator (Chapter IX)—is situated at the foot of the aerial and supplies energy to the latter by means of some form of inductive or capacitive coupling (fig. 16a). The distribution of current and voltage in such an aerial are shown in fig. 16b. If the earth is assumed to be perfectly conductive the distribution of the electric field around the aerial is exactly the same as round the upper half of a dipole, and the earthed vertical wire may be considered to form one half of a dipole, the other half of which is buried in the earth. The electric fields now terminate upon the surface of a theoretically perfect conductor instead of forming closed loops in free space, and consequently high frequency currents circulate in the earth between regions of positive and negative electrical sign. As in reality the earth is by no means a perfect conductor, the distribution of the field is not exactly as has been described, but a more important effect of its finite conductivity is that the earth currents give rise to a loss

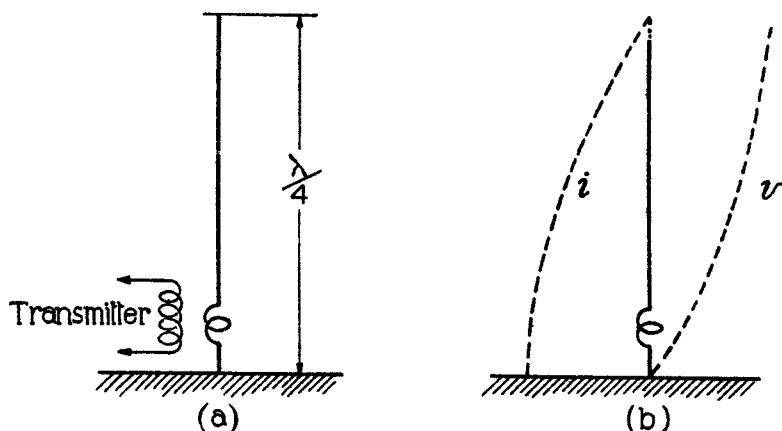


FIG. 16, CHAP. VII.—Current and voltage distribution, quarter-wave aerial.

of energy, the rate at which energy is expended being proportional to the square of the aerial current and to the specific resistance of the earth in the immediate vicinity of the station. A large counterpoise of low resistance, some six feet above the ground, will reduce these losses considerably and such a counterpoise when used in preference to an actual earth connection is called an earth screen, because its function is to prevent the flow of radio-frequency currents in the earth underneath the aerial. It should preferably extend a distance at least equal to one half the height of the aerial in all directions, but need not be a complete metallic area, a system of radial wires spaced 15 or 30 degrees apart being equally efficient. Even if the ideal earth screen cannot be adopted, a few wires arranged in this way will often reduce the energy wastage considerably.

L and T aërials

47. In order to increase the capacitance between aerial wire and earth or counterpoise, a conducting area is frequently added to the upper extremity. The advantages of this addition are twofold. First, the increase of capacitance allows a greater storage of energy for a given voltage and therefore a larger charging current in the vertical portion, and second, the current distribution in the vertical portion is more nearly uniform throughout its length. The significance of the latter will be appreciated when the "effective height" of the aerial has been defined. The portion

added usually consists of horizontal members, giving rise to the forms known as L or T-shaped aeriels, which are shown in figs. 17 and 18. Where only a single mast is available, the so-called "umbrella" type may be employed, although its use has been largely discontinued owing to its rather poor radiating properties, while if more than two masts are permissible, a triangular or rectangular network may be employed. Radiation takes place not only from the vertical portion (usually called the "feeder") but also from the upper capacitance area (or "roof") but the radiation from the latter is not usually coincident in direction and polarisation with that from the former. The current and voltage distribution of L and T aeriels is similar to that of a vertical aerial of the same overall length, measured from the base of the aerial to the extremity. The T aerial may be considered to consist of two L aeriels placed "back to back", and the current and voltage distribution of both types are shown in fig. 19. The distribution in other forms of aerial having added capacitance areas is similar to these, but do not lend themselves to diagrammatic representation because the aerial occupies three spatial dimensions.

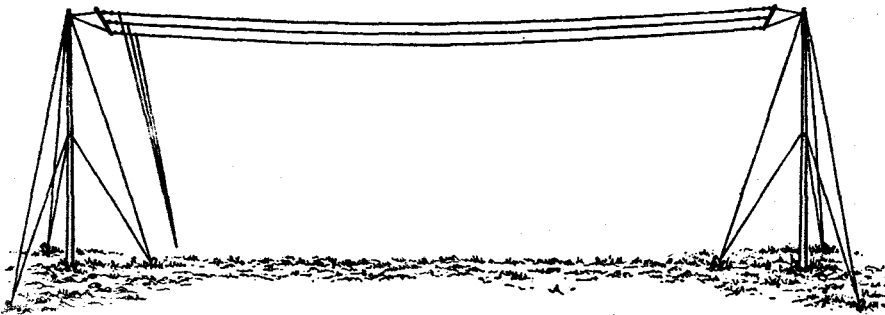


FIG. 17, CHAP. VII.—L-aerial.

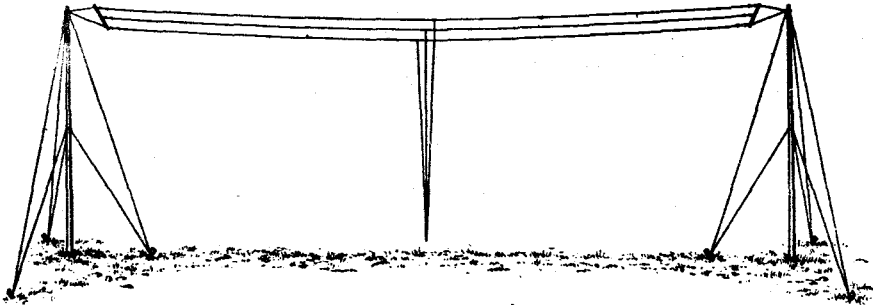


FIG. 18, CHAP. VII.—T-aerial.

Aerial tuning

48. Under certain conditions it is desirable to employ a single aerial to transmit any one of a large number of frequencies, e.g. the frequency found to be most effective for communication between aircraft and a ground station may not be equally effective for intercommunication between aircraft or between ground stations. This introduces the notion of artificially lengthening or shortening the aerial wire. Artificial lengthening, in order to decrease the resonant frequency, can be achieved by adding wire in the form of a coil, i.e. an inductance, at the most convenient point, which is where the transmitter is located, while if a condenser is similarly inserted the resonant frequency will be increased. The resonant frequency of the aerial circuit can be decreased to any extent by the introduction of a sufficiently large inductance, but it is important

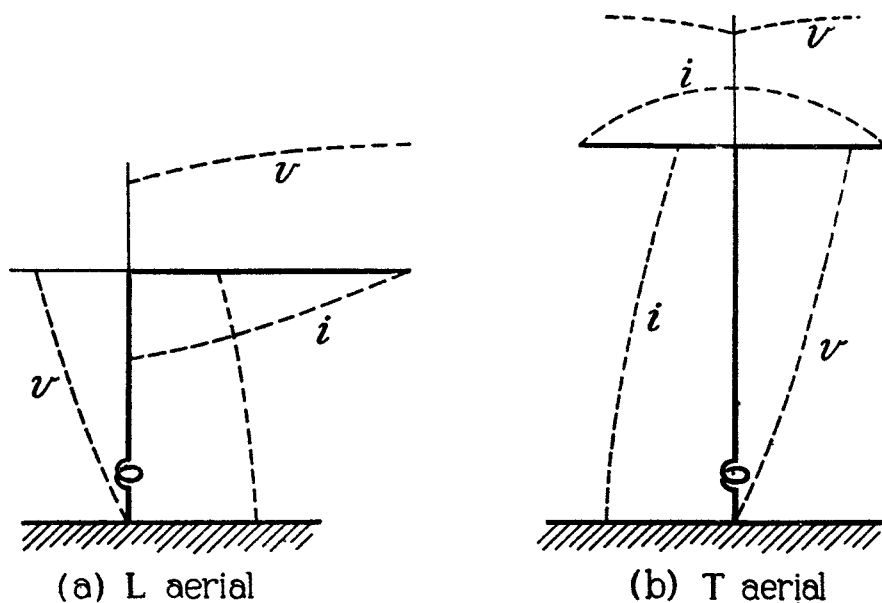


FIG. 19, CHAP. VII.—Current and voltage distribution, L and T aërials.

to remember that even the interposition of an infinitely small condenser, i.e. of zero capacitance, will tune the aërial to a frequency only twice the natural frequency, for a capacitance of zero value is equivalent to a complete break in the circuit and the aërial then becomes an isolated wire, i.e. a dipole or half-wave aërial. The current and voltage distribution is modified by the

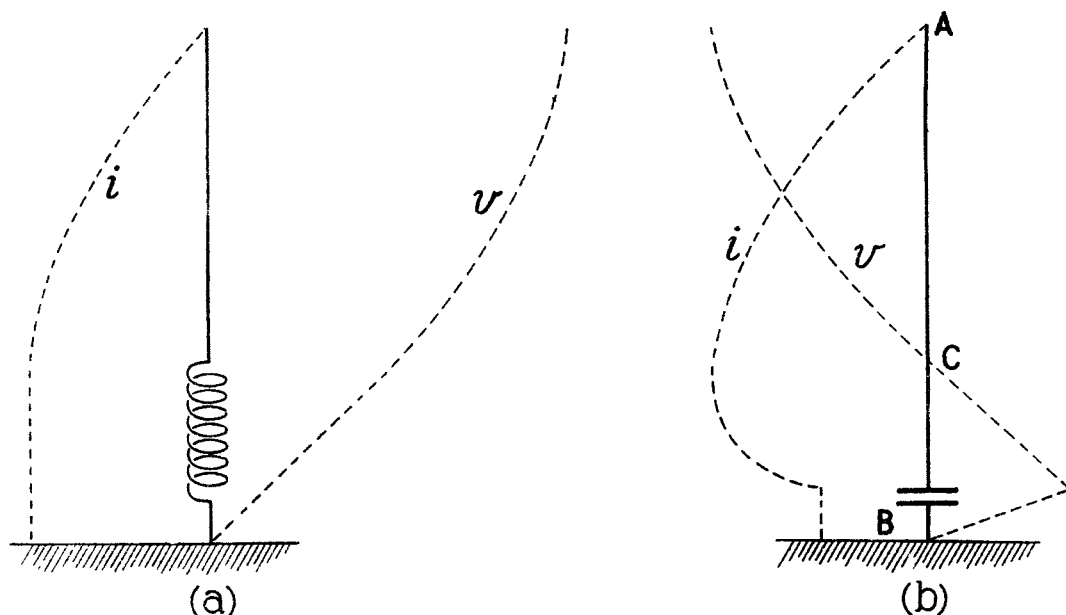


FIG. 20, CHAP. VII.—Current and voltage distribution of vertical wire with added reactance.

insertion of these tuning devices. Consider the most usual case, which is the addition of inductance. The coil should have very small capacitance with respect to the earth or counterpoise, and little displacement current should occur in its vicinity, hence the current at both ends should be nearly the same. A considerable increase of voltage will take place between the same points, hence the current and voltage distribution will be somewhat as shown in fig. 20a. Let us now consider the effect of a series condenser. In fig. 20b, AB represents the aerial, and without added reactance it would be a quarter-wave aerial having the current distribution shown in fig. 16b. The insertion of the condenser reduces its effective length to AC (say), and therefore a current loop must exist at the point C, while below this point the current decreases. The current distribution is therefore that shown by the curve and the voltage distribution can be deduced from the current distribution, since a voltage loop must exist at the upper extremity and voltage nodes at the point C and at the earth connection. The current distribution diagram illustrates that when a series condenser is in use, the reading of a thermoammeter in the wire immediately adjacent to the earth connection may be a very unreliable indication of the radiating properties of the aerial.

Distribution of radiation in space

49. (i) Returning to a consideration of the vertical dipole situated far above the earth's surface, we may define the equatorial plane as an imaginary plane passing through the midpoint of the dipole and perpendicular to it. The point at which the dipole passes through this imaginary plane will be called the pole. At all points on this plane which are equidistant from the pole the radiation field will be of equal strength. This is not true for all forms of aerial, and it is usual to indicate the directive radiating properties of any particular type of aerial by what are called polar diagrams. The horizontal polar diagram shows the field strength in various directions in the equatorial plane, and theoretically is constructed as follows. From the pole are drawn straight lines on the equatorial plane, radiating in all directions, the length of each line being proportional to the field strength (or sometimes which is proportional to the square of the field strength) in that particular direction. The ends of these lines form the horizontal polar diagram. As the dipole radiates equally well in all directions in the equatorial plane, which is apparent from consideration of its symmetrical arrangement in space, the polar diagram will be a circle having the pole as its centre.

(ii) The horizontal polar diagram of a vertical earthed aerial, the height of which is one quarter of a wavelength or less is also a circle having the aerial as its centre, the equatorial plane in this instance being practically coincident with the surface of the earth. In practice, the polar diagram of an aerial situated near the earth's surface is obtained by a method which is essentially as follows. Upon a suitable map is described a circle of given radius, say 100 miles, with the location of the aerial as its centre. A special form of radio receiver is carried from point to point round the circumference of this circle, and the strength of the radiation from the transmitter is measured at each point by means of this receiver. These measured "signal strengths" are then drawn to a convenient scale along the radii connecting the pole with the point at which each measurement was made, measuring outward from the pole, and the line joining the ends of these radii forms the polar diagram. It is obvious that the actual measurement takes considerable time, and the conditions at the transmitter must be kept constant during the whole of the time during which measurements are actually being undertaken. Fig. 21 shows a number of horizontal polar diagrams of a B.B.C. transmitter situated in London, each diagram being marked with the electric field strength, in millivolts per metre, the whole forming what is called a field strength contour diagram. The aerial in this instance was of the "T" type and was erected upon steel masts. The line joining these masts, if produced, was found to coincide, approximately, with the directions in which the field strength was least, i.e. roughly in the N.E. and S.W. directions. The marked reduction in field strength in these directions may therefore be attributed to losses in the steel masts. When, however, the masts are so designed that such losses are unimportant, it is sometimes found that an L or T aerial possesses directional properties, but these may not be very apparent unless the length of the horizontal portion is several times the vertical height, the field strength being greatest in the direction of

CHAPTER VII.—PARA. 50

the flat top. As a rough guide it may be assumed that no appreciable directional effect will be attained unless the ratio $\frac{\text{length}}{\text{height}}$ exceeds ten, and therefore when erecting an aerial of this kind for a given service there is nothing to be gained by siting the masts in any particular direction, unless the above ratio is fulfilled. Special aerial designs for directional transmission will be discussed in a later chapter.

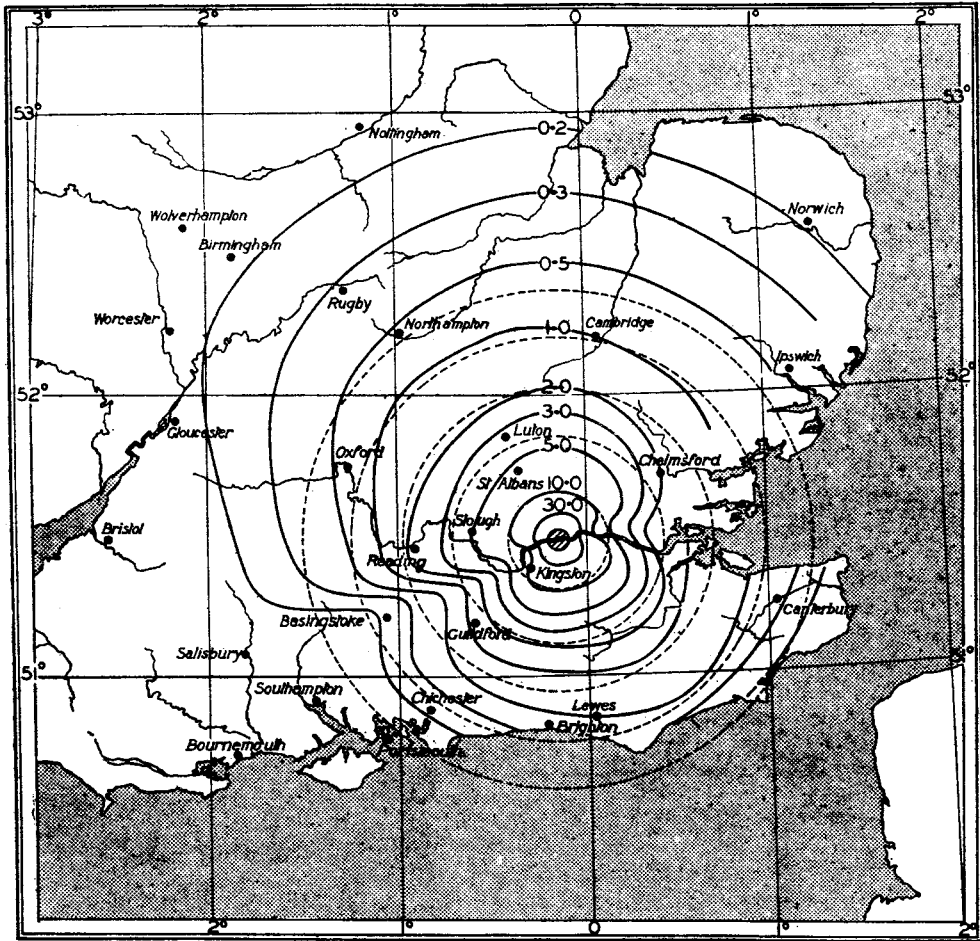


FIG. 21, CHAP. VII.—Field strength contour diagram.

Vertical polar diagram

50. (i) This diagram shows the distribution of field strength in any vertical plane through the aerial. Commencing as before with the dipole in free space the radiation fields are of maximum intensity in a direction perpendicular to the dipole, and of zero intensity along its axis. The vertical polar diagram consists of two circles, and is given in fig. 22a, while a quarter-wave aerial, situated upon a perfectly conducting earth, would have radiating properties identical with the upper half of this (fig. 22b).

(ii) When an aerial is situated near to the ground, its directional properties in the vertical plane are affected by the proximity of the earth, because the energy radiated towards the earth is reflected, and the reflected wave will either reinforce or detract from the energy received at points above ground level. As the earth is not a perfect conductor, some absorption of energy also occurs but no definite allowance can be made for this since the absorption depends upon local conditions. The vertical polar diagrams of a vertical half-wave aerial, situated at various distances above a perfectly conducting earth are shown in fig. 23 by the full lines, and the effect of finite conductivity of the surrounding soil is somewhat as shown by the dotted outline.

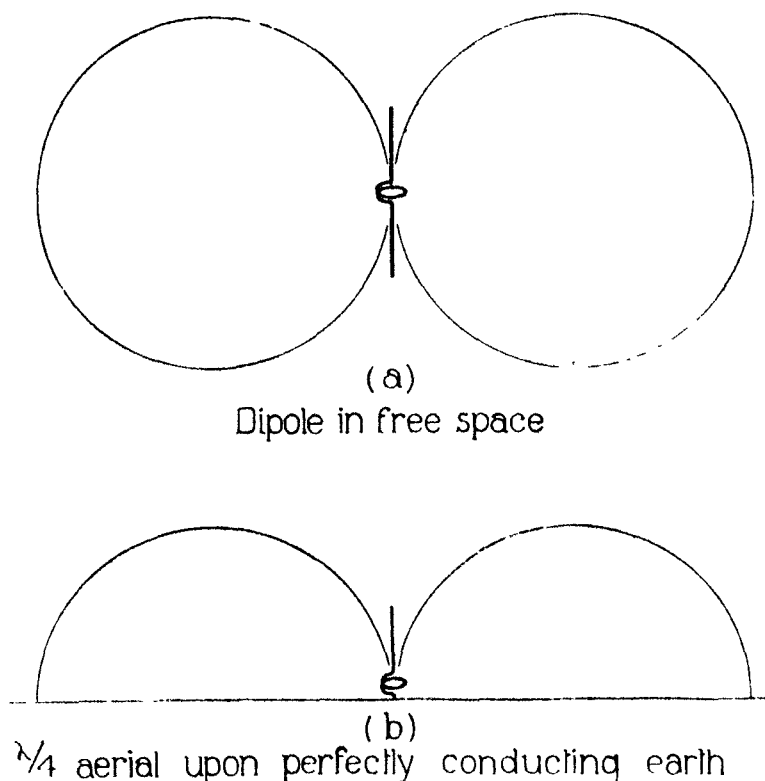


FIG. 22, CHAP. VII.—Vertical polar diagrams.

Radiation resistance and aerial efficiency

51. Of the total energy supplied to a transmitting aerial a portion is transferred to the ether and is radiated into space, while the remainder is dissipated in the form of heat either in the aerial itself or in its immediate surroundings. The rate at which energy is supplied to the aerial, or input power P , must be capable of expression in the form I^2R , because the latter is a perfectly general form, but the terms current and resistance require special definition when used in connection with aërials. The current distribution is never uniform, and it is therefore usual to refer to the value of the current at the current loop nearest to the feeding point as the aerial current. The aerial resistance is then defined as that quantity which, when multiplied

CHAPTER VII.—PARA. 51

by the square of the aerial current, gives the power supplied to the aerial circuit. Alternatively it may be stated that the aerial resistance R_A is defined by the equation

$$I_A^2 R_A = \text{input power}$$

where I_A is the R.M.S. value of the current at the current loop nearest the feeding point. The value of resistance thus defined is called the total effective resistance of the aerial, and may be divided into two components corresponding to (i) the power dissipated in the form of heat in the aerial and its surroundings, which may be written $I_A^2 R_h$ watts, R_h being the loss resistance of the aerial, and (ii) the power converted into radiation which may be written $I_A^2 R_r$, R_r being termed the radiation resistance of the aerial. This radiation resistance must be regarded as purely fictitious, and may be defined as the imaginary ohmic resistance which if supplied with a certain current would convert energy into heat at the rate at which energy is converted into radiation by the aerial when supplied with an equal current. The total effective resistance of the aerial is equal to the sum of the radiation resistance R_r and the loss resistance R_h , or $R_A = R_h + R_r$.

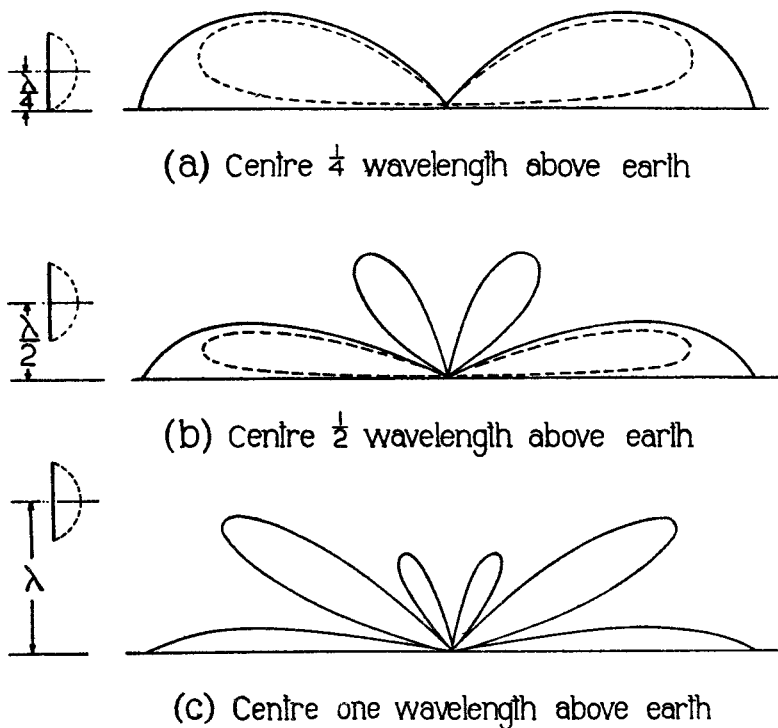


FIG. 23, CHAP. VII.—Vertical polar diagrams of half-wave aerial showing effect of elevation of aerial.

The efficiency of an aerial as a radiator can be expressed in the same manner as for all other electrical machinery, namely, efficiency = $\frac{\text{Power converted into useful work}}{\text{Total power supplied}}$ hence, if efficiency is denoted as usual by η

$$\begin{aligned} \eta &= \frac{I_A^2 R_r}{I_A^2 (R_h + R_r)} \\ &= \frac{R_r}{R_h + R_r} \end{aligned}$$

52. (i) The radiation resistance of an aerial depends upon its shape and its dimensions relative to the wavelength of the radiation emitted, while its surroundings affect both the radiation resistance and the loss resistance. It is possible to derive a theoretical expression for the radiation resistance on the assumption that the capacitance of the aerial is concentrated between the extreme ends of the wire, which is tantamount to the assumption that the current distribution is uniform over the whole length of the aerial. The imaginary radiator having this current distribution is shown in fig. 24. It is called a hertzian doublet, and must not be confused with the dipole or half-wave aerial in which it will be remembered a current loop exists at the middle point while a current node is found at each end. The radiation resistance of such a doublet in free space is given by the expression

$$R_r = \frac{80 \pi^2 l^2}{\lambda^2}$$

where l is the length of the doublet.

For example, if the length of the doublet were $\frac{\lambda}{2}$

$$R_r = \frac{80 \pi^2}{\lambda^2} \times \frac{\lambda^2}{4} = 197 \text{ ohms.}$$

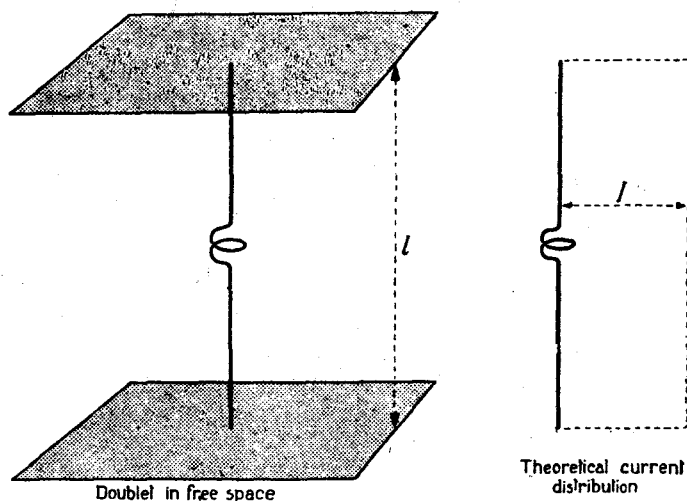


FIG. 24, CHAP. VII.—Hertzian doublet showing assumed current distribution.

(ii) The dipole differs from the doublet in that the current distribution is co-sinusoidal, being a maximum at the mid-point and varying in such a way that if I is the current at the middle of the dipole and i the current at a point x centimetres from the latter point, $i = I \cos \frac{2\pi}{\lambda} x$. At the end of the wire $x = \frac{\lambda}{4}$ and $i = I \cos \frac{\pi}{2} = 0$. The average value of the current over the whole length of wire is $\frac{2}{\pi} I$ and the dipole in free space will have a radiation resistance equal to that of a doublet of the same length but in which the current is only $\frac{2}{\pi} I$, or alternatively a

CHAPTER VII.—PARA. 52

doublet in which the current is I amperes throughout its length, but with the latter dimension reduced by a factor $\frac{2}{\pi}$. Hence the length of the doublet equivalent to a half-wave aerial is $\frac{\lambda}{\pi}$, and the radiation resistance is

$$R_r = \frac{80 \pi^2}{\lambda^2} \times \left(\frac{\lambda}{\pi}\right)^2 = 80 \text{ ohms.}$$

This is the theoretical radiation resistance of the dipole in free space. Similarly the radiation resistance of a quarter-wave aerial earthed at its lower end is one-half that of the dipole in free space, i.e. 40 ohms. More accurate calculations show that the radiation resistance is actually rather less than the value derived from these formulae, which is partly due to the fact that the fundamental wavelength of a so-called quarter-wave aerial is not exactly four times its length, but is about 4.18 times. This "end correction" may be regarded in this way. Although it has been stated that no conduction current can flow past the end of the aerial wire, the variations

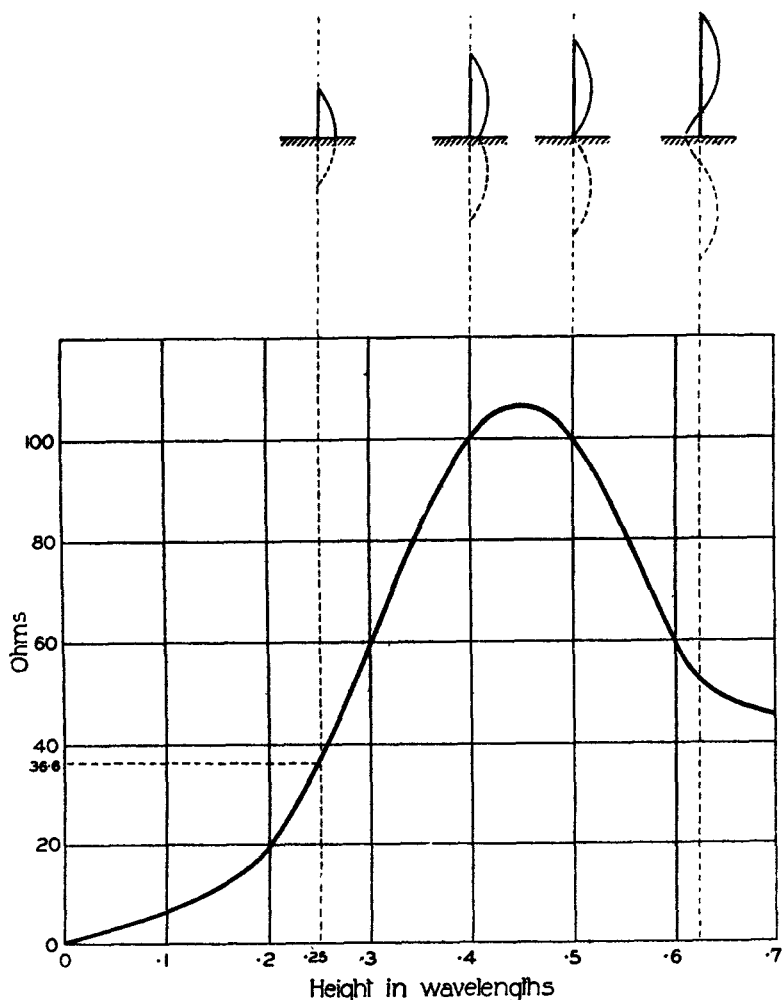


FIG. 25, CHAP. VII.—Radiation resistance of vertical aerials of height less than $.7\lambda$.

of its potential cause displacement current to flow at this point, and some lines of electric strain must exist which are, geometrically, merely continuations of the lines of moving electrons constituting the conduction current. Hence the aerial wire may be regarded as being continued for a short distance into the dielectric. Applying a correction for this end effect the above resistances become for a dipole in free space $80 \times \left(\frac{4}{4 \cdot 18}\right)^2 = 73 \cdot 25$ ohms and for the earthed quarter-wave aerial $36 \cdot 6$ ohms. The radiation resistance of a vertical half-wave aerial, situated near the earth, is approximately equal to one half that of a doublet in free space, namely 98 ohms approximately.

(iii) The radiation resistance of vertical aeriels having a height up to $0 \cdot 7 \lambda$, the lower end being at or near ground level, is shown graphically in fig. 25, in which, for heights below $\cdot 25 \lambda$, the power radiated is the product of the given radiation resistance and the square of the current at the base of the aerial, while for aeriels having a height greater than $\cdot 25 \lambda$, the power radiated is the product of the given radiation resistance and the square of the current at the current loop. The radiation resistance is said to be calculated with reference to base current or to loop current respectively.

53. The radiation resistance of L or T aeriels less than a quarter of a wavelength long (i.e. those used for the medium and lower radio frequencies), is usually computed by employing a formula derived from the conception of the doublet, but introducing the idea of "effective height". For any given ratio of length of roof to height of an L aerial, a corresponding "form factor" is employed. The effective height h_e is then defined as the true height h_o of the aerial above the ground, multiplied by the form factor F corresponding with the ratio of length to height for that particular aerial. The effective height is then inserted in the formula

$$R_r = \frac{160 \pi^2 h_e^2}{\lambda^2}$$

which is derived from the doublet theory, the form factor being intended to correct for the distribution of current along the aerial. The value $\frac{2}{\pi}$ already used for this purpose is in fact the form factor for an earthed vertical aerial, for which the ratio $\frac{\text{length of roof}}{\text{vertical height}} = 0$. The form factor may be calculated from the empirical formula

$$F = \frac{2 + 8 \frac{l}{h_o}}{\pi + 8 \frac{l}{h_o}}$$

where l is the length of the horizontal portion. This expression gives results within about 2 per cent. of the theoretical value.

Losses in aeriels

54. The loss resistance may be divided into four components, namely:

- (i) conductor losses in the aerial system, including the earth or counterpoise,
- (ii) eddy current losses in adjacent conductors,
- (iii) losses due to imperfect insulation,
- (iv) dielectric losses.

Conductor losses are inherent in any electric circuit, but can be maintained at a minimum by attention to those factors which cause an increase of resistance at high frequencies. It is, however, always necessary in practice to compromise between electrical efficiency, mechanical robustness and cost. At fixed ground stations it is possible to reduce the conductor losses to a minimum by the use of high conductivity copper for aerial conductors, by employing several

CHAPTER VII.—PARA. 54

wires in parallel in the aerial, feeder and earth wires, and by utilising conductors of separately insulated stranded wire (litzendraht) for the tuning inductances. A point frequently overlooked is that the current distribution along the length of the aerial is not uniform, and it is therefore preferable that the current-carrying capacity should be greater in the vicinity of the current loop or loops than elsewhere. It is not usually feasible to arrange this except in the case of the lower radio frequencies, in which an earthed L-aerial is operated at a frequency below its fundamental. Under these conditions the number of wires used in the vertical portion or feeder should be at least equal to and preferably greater than the number of wires used in the flat top. This is hardly practicable when a T-aerial is employed owing to the increased sag set up in the "roof" by the additional weight of the feeder, tending to mechanical weakness and reducing the effective height. Another point of importance is the connection point of the feeder to the flat top. Unless it is possible to connect the feeder at the electrical centre of the span, and to lead off at right angles to the latter, the L-aerial will generally be found superior to the T-aerial. Finally, if it is necessary to join wires, such joints must be soldered, using as little heat as possible consistent with efficient soldering, and making the joint as quickly as possible. It should hardly

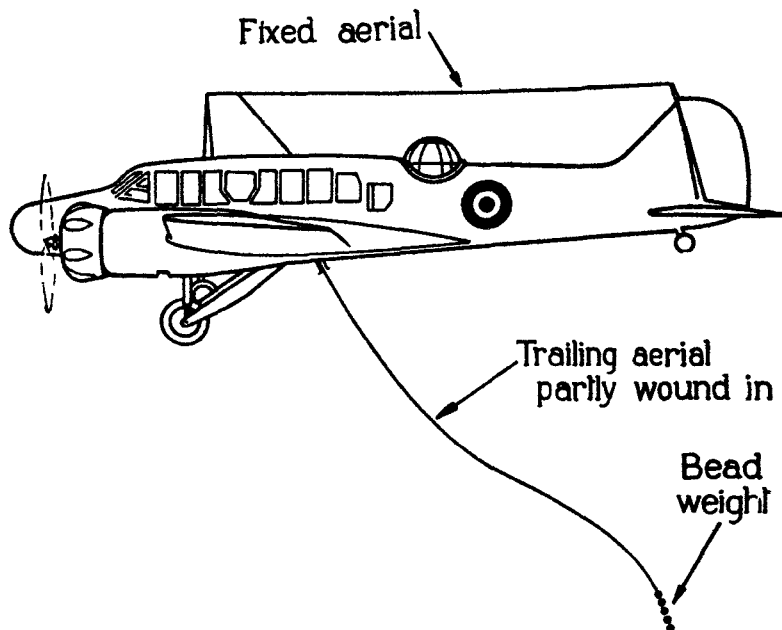


FIG. 26, CHAP. VII.—Arrangement of aerals on aeroplane.

be necessary to point out that where a specification is laid down for the erection of an aerial, it should be followed implicitly. In aircraft, it is often necessary to accept a type of aerial which is electrically far from ideal. Thus where the frequency of transmission renders a long aerial necessary, the only possible form is a trailing wire, fitted with a weight at its extremity. Either phosphor-bronze or stainless steel may be employed in different circumstances, the former combining moderate mechanical strength with fairly high conductivity, and the latter high mechanical strength and durability with rather lower conductivity at radio frequencies. At the higher radio frequencies used by aircraft it is possible to utilise fixed aerals which are erected upon the airframe according to specifications laid down for the particular aircraft. A typical example is shown in fig. 26. In aircraft, where the whole metalwork of the machine is utilised as a counterpoise, the bonding of these parts must be maintained at its highest possible efficiency. Conductor losses in the aerial and earth or counterpoise may be considered to vary as the square root of the frequency.

Eddy current losses in adjacent conductors are caused by induced currents in the latter. In ground stations, masts and stays are the principal causes of loss (cf. paragraph 49 and fig. 21) which is minimised by either ensuring that the masts and stays are well earthed, or alternatively highly insulated from earth. Eddy current losses vary with the square root of the frequency as do other conductor losses.

Losses due to imperfect insulation are reduced to a minimum by the employment of insulators of high efficiency and dielectric strength, such as glass or porcelain. These properties are very much impaired by atmospheric pollution, and the insulation can only be maintained at its highest possible value by frequent cleaning. While this process may be difficult in large ground stations where traffic is almost continuously handled, it is possible in aircraft to perform an inspection of all aerial insulation at frequent intervals, and to maintain this insulation at a high standard. Dielectric losses in the aerial may be assumed to vary inversely as the frequency.

Dielectric losses in adjacent insulating materials such as masonry, wood, trees, etc., are caused by dielectric hysteresis. The less efficient the dielectric, the higher these losses will be. They may be considered to vary inversely as the frequency.

55. The manner in which these components of aerial resistance vary with frequency may be exhibited in graphical form as in fig. 27 in which aerial resistance is plotted against frequency. Curve (i) shows the radiation resistance, and its shape indicates that this quantity varies directly as the square of the frequency. Curve (ii) depicts the nature of variation of the conductor losses, including those caused by eddy currents, and these are proportional to the square root of the frequency. Curve (iii) shows the manner in which the dielectric losses vary; these are inversely proportional to the frequency as previously stated. The curve (iv) showing the total loss resistance is derived from these, by adding the value of curves (ii) and (iii) for several values of frequency and plotting this sum to give curve (iv). It will be observed that for any given aerial

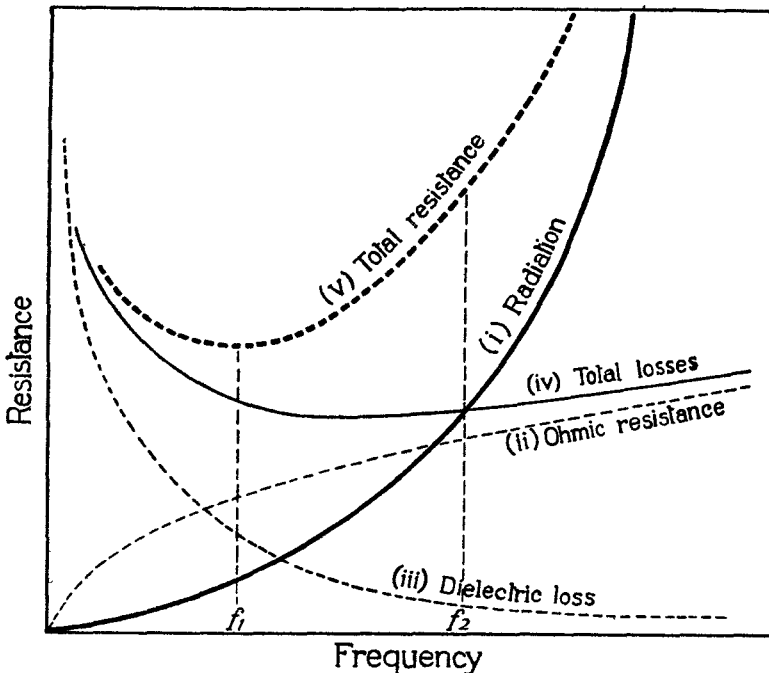


FIG. 27, CHAP. VII.—Components of aerial resistance.

CHAPTER VII.—PARA. 56

there is always a particular frequency at which the total loss resistance is a minimum. The total effective resistance is given by curve (v) which is obtained by adding curve (i) (radiation resistance) to curve (iv) (total loss resistance). This curve also has a minimum value, showing that at one particular frequency a given aerial current can be obtained with smaller input power than at any other frequency. This is not the most economical frequency of operation however, for the efficiency under these conditions is low. In the diagram, the frequency at which the total effective resistance is least is denoted by f_1 , and the radiation resistance is only about one-fifth of the total resistance, giving an aerial efficiency of 20 per cent. At the frequency denoted by f_2 , however, the radiation resistance has increased to such an extent that it forms one-half of the total resistance giving an efficiency of 50 per cent.

High frequency aerals

56. It is now proposed to consider the principles governing the design of non-directive aerals for high frequencies, i.e. those lying in the band between 3 and 30 megacycles per second, the corresponding wave-lengths being 100 to 10 metres. It has hitherto been assumed that it is

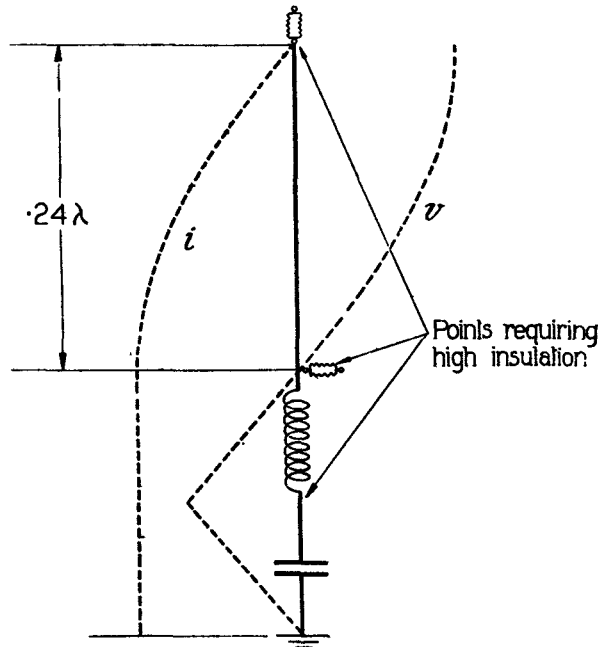


FIG. 28, CHAP. VII.—Quarter-wave aerial terminated by acceptor circuit.

desirable to erect a transmitting aerial in such a manner that radiation is normally polarised at the moment of its inception. Owing to the mechanism by which waves of high frequency are propagated, however, it is possible to utilise either vertical or horizontal radiating elements, at any rate for long distance communication. The height of the aerial, compared with the wave-length, is of great importance. If a horizontal dipole is erected very near to the earth's surface, the effect of reflection by the ground is to cause most of the energy to be radiated upwards, the vertical polar diagram being roughly a circle touching the ground at the point immediately beneath the aerial. To obtain appreciable radiation at an angle of 30° to the ground, it is necessary to elevate a horizontal radiator to a height of at least one half-wavelength above the ground. At the lower frequencies of the band under discussion, it may be necessary

for extempore construction to accept a vertical quarter-wave aerial, e.g. for operation on a wavelength of 100 metres, the height of the aerial would be $\cdot 24\lambda = 24$ metres or 73 feet. The total aerial resistance will then be of the order of 100 ohms, and energy may be transferred to the aerial by connecting its lower end to an acceptor circuit tuned to the same frequency, the mode of connection and the current and voltage distribution being given in fig. 28.

High frequency feeder line

57. A half-wave aerial for this frequency would have an actual length of about 146 feet, and it will rarely be practicable to erect a vertical mast of such a height from service resources alone, but within the limitations suggested in the preceding paragraph, a horizontal half-wave aerial on 70-ft. masts is quite a feasible proposition, and again the method of transferring energy to the aerial of the greatest importance. As the aerial is about a quarter wavelength from the ground, where the transmitter must be situated, some form of feeder line must be employed. High frequency feeders may take one of two forms, namely the resonant and non-resonant types. Various forms of construction are possible, but the only kind of line which lends itself to extempore design and erection is that consisting of twin parallel wires, mounted upon poles if necessary, and well insulated from earth. It is generally most important that the two wires shall be absolutely symmetrically disposed with regard to the aerial, the ground beneath the aerial and the transmitter.

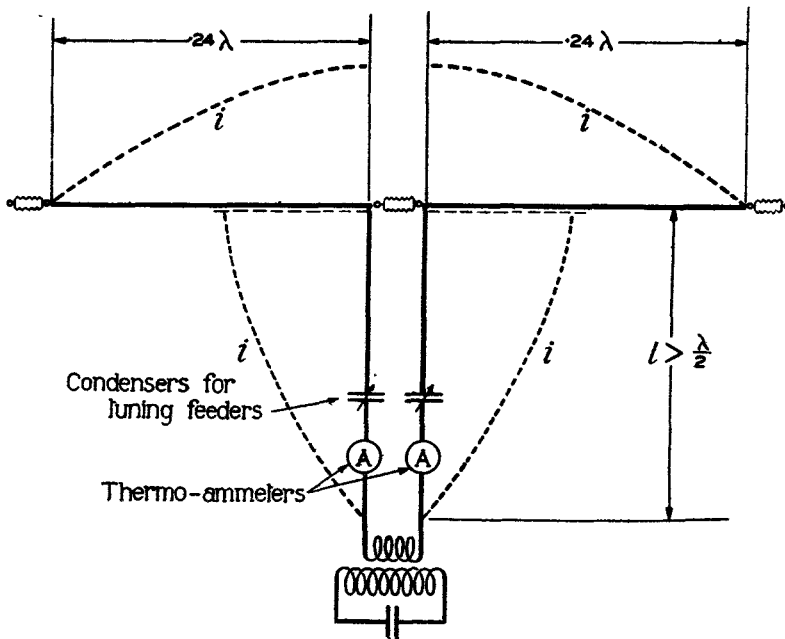


FIG. 29, CHAP. VII.—Current-fed half-wave aerial with tuned feeder.

Feeding by resonant line

58. The resonant line is only suitable for use when the distance between aerial and transmitter is not much greater than $\frac{\lambda}{2}$. This limitation arises because (i) a stationary wave is set up in each feeder and the peak value of oscillatory current rises to a value equal to that in the aerial; the ohmic losses are therefore high. (ii) It is difficult to ensure absolute symmetry

CHAPTER VII.—PARA. 59

between the two feeders and this gives rise to radiation from the feeder, again leading to excessive power loss, for this radiation will not generally be in exact phase with that from the aerial and therefore will rarely act merely as a reinforcement of the latter. Assuming that the aerial is erected sufficiently near to the transmitter to admit the employment of a resonant feeder, the next point to be decided is whether it is desirable to feed at a current loop or voltage loop. This is merely a matter of practical convenience, and no electrical superiority exists in either method. If current loop feeding is adopted, the half-wave aerial is divided at its centre by a glass or porcelain strain insulator, each half being connected to one feeder. The electrical length of the feeders should be $\frac{\lambda}{2}$, i.e. either their actual length slightly greater than this, their effective length being adjusted by variable condensers connected in series with each line, or alternatively slightly less than $\frac{\lambda}{2}$ their effective length being increased by an adjustable inductance in each line. A

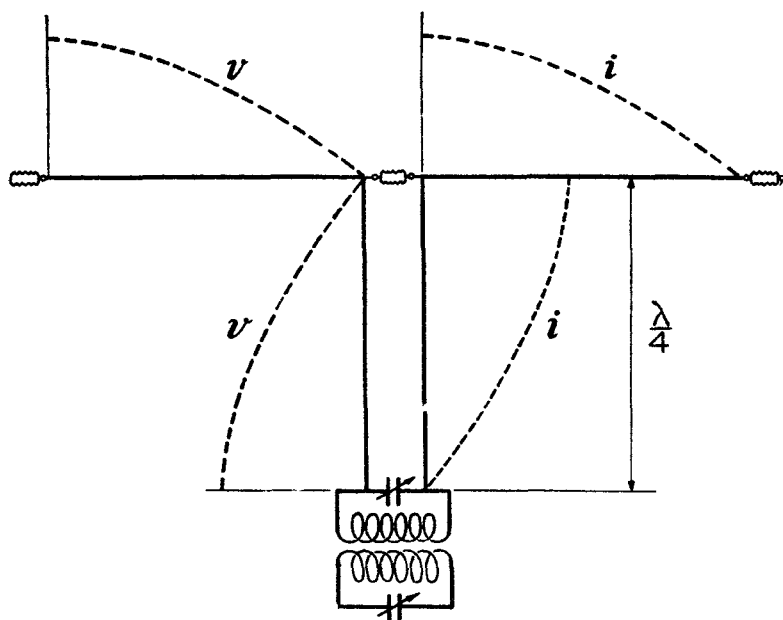


FIG. 30, CHAP. VII.—Current-fed half-wave aerial with $\frac{\lambda}{4}$ feeder.

thermoammeter in series with each will also enable a current balance to be achieved, and the feeder may be considered to be resonant when the sum of the two feed currents is a maximum. The effective length of the coupling coil by which energy is transferred from transmitter to line must also be taken into account. The current distribution of an aerial fed in this way is shown in fig. 29.

59. If the aerial and transmitter are so close that a $\frac{\lambda}{4}$ resonant feeder is preferable the desired current and voltage distribution can be achieved by inserting a rejector circuit, tuned to the desired frequency, in place of the coupling coil mentioned above. The circuit then becomes that given in fig. 30, and it will be observed that the rejector circuit replaces a length of feeder equivalent to one half-wavelength. Again, the relative positions of transmitter and aerial may render it more convenient to feed the half-wave aerial at one end. Now in the current feed arrangement, each half of the aerial is effectively a $\frac{\lambda}{4}$ aerial, having a radiation resistance of

about 36 ohms or 72 ohms in all, and the loss resistance may amount to a further 20 ohms in each half, so that the input resistance of each is in round figures 100 ohms. It will be noted that this figure is attained when feeding at a point where the current is large and the voltage small, i.e. the ratio $\frac{V}{I}$ is small (100 ohms). When feeding at the end of the $\frac{\lambda}{2}$ aerial, however, the current

being a minimum and the voltage a maximum, the ratio $\frac{V}{I}$ is large, being equivalent to about 2,500 to 3,500 ohms. It may be stated that whereas the load imposed upon the source by an aerial fed at a current loop resembles that of an acceptor circuit, the load imposed when fed at a voltage loop resembles that of a rejector circuit. The figure 2,500 to 3,500 ohms above mentioned is in fact calculated on the assumption that the average inductance of such an aerial is about $1.85 \mu H$ per metre and its capacitance about $.000006 \mu F$ per metre, hence the dynamic resistance of the aerial is :

$$\begin{aligned}\frac{L}{CR} &= \frac{1.85 \times \frac{\lambda}{2}}{.000006 \times \frac{\lambda}{2} \times (72 + 40)} \\ &= \frac{1.85 \times 10^6}{6 \times 112} \\ &= 2,750 \text{ ohms.}\end{aligned}$$

In general it may be said that when a $\frac{\lambda}{2}$ aerial is fed at the end, the circuit by which energy is supplied should have the characteristics of a rejector circuit, while if fed at the centre, which is a current loop, the feeder should have the characteristics of an acceptor circuit. In this way the maximum transference of energy is necessarily obtained, and the feeders are said to be "matched" to the aerial. The importance of matching the various parts of the circuit is of even greater importance when a non-resonant line is employed.

60. One possible method of arranging the feeder is shown in fig. 31. It will be seen that although a twin-wire feeder is employed, only one of the wires is actually in use, the other being

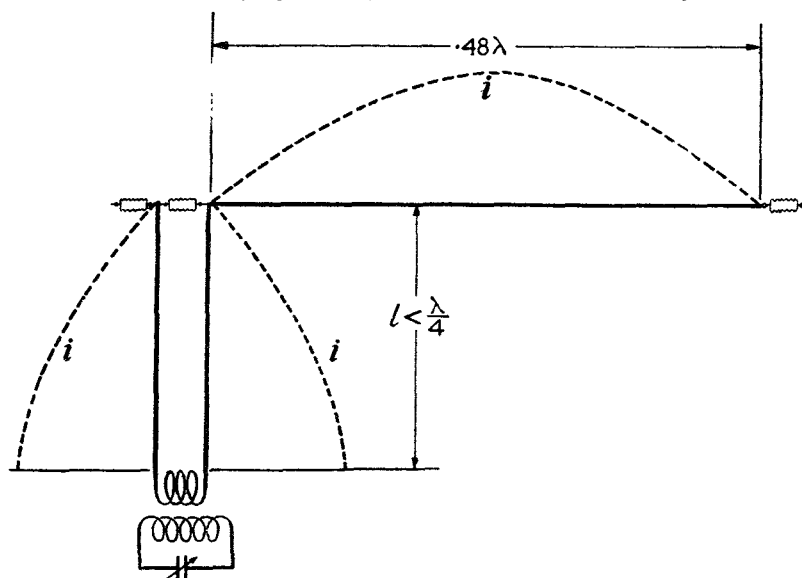


FIG. 31, CHAP. VII.—Voltage fed half-wave aerial.

CHAPTER VII.—PARA. 61

used merely in an attempt to eliminate radiation from the feeder. This elimination cannot be complete because the current at the end of the inactive feeder is necessarily zero, but that at the end of the active feeder is the feed current of the aerial and will be of the order of one-fifth that at the current loop. The feeders themselves are approximately $\frac{\lambda}{4}$ in length, and are made electrically equivalent to $\frac{\lambda}{4}$ by adding either inductance or capacitance in each line at the transmitter end. A simple coupling coil serves to transfer energy from transmitter to feeder. The foregoing discussion is of course equally applicable to the higher frequencies of the 3—30 Mc/s band, half-wave or three-quarter-wave vertical aerals being usually employed. The latter form of aerial gives maximum radiation in a direction inclined about 30° to the horizontal, and this may be advantageous for long distance transmission in that less energy is absorbed by the earth in the neighbourhood of the transmitter. $\frac{3}{4}\lambda$ aerals are fed in the same way as $\frac{\lambda}{4}$ aerals.

Non-resonant transmission lines

61. The theory of the transmission line will receive more detailed consideration in Chapter XV to which the following may be considered an introduction. It has already been pointed out that any long length of conductor, an aerial being an example, has inductance, capacitance to earth and to adjacent conductors, and also resistance, distributed throughout its length, thus the resonant feeders and aerals previously considered may be regarded as transmission lines of such a length that the waves reflected from the ends of the conductor reinforce or cancel each other at certain points. If L' is the inductance and C' the capacitance, both measured in C.G.S. units per centimetre, the velocity with which an electromagnetic disturbance travels along the conductor is practically equal to the velocity in free space, namely 3×10^{10} centimetres per second, and the latter is also equal to $\frac{1}{\sqrt{L'C'}}$. Suppose that a twin wire transmission line is supplied with an alternating E.M.F. of any frequency whatever, and that the line is of infinite length, so that no reflected wave can occur. This absence of a reflected wave implies that the line is non-resonant, and behaves in the same way at all frequencies. Assuming that the inductance and capacitance are equally distributed along the whole length, and that the resistance is negligible, the current in the line will decrease gradually along the line owing to the displacement current between lines, hence the line possesses an effective impedance, called its characteristic or surge impedance, which can be shown to be equal to $\sqrt{\frac{L'}{C'}}$ ohms. The current and voltage in such a line are in phase with each other at all points and therefore the surge impedance is actually an effective resistance. It must be noted that as L' and C' are the inductance and capacitance per centimetre, any length of the same line has the same surge impedance, namely $\sqrt{\frac{L'}{C'}}$ ohms. Now in practice we cannot have a line of infinite length; if, instead, the line is of finite length, and the ends of the wires remote from the generator are connected to a non-inductive resistance equal to $\sqrt{\frac{L'}{C'}}$ ohms, the load imposed upon the generator by the finite line and its terminating resistance is exactly the same as the load imposed by an infinitely long line, for the current and voltage are in phase at all points and no stationary wave is set up. The physical meaning of this is that energy is converted into heat in the terminating resistance at exactly the rate at which it arrives at the end of the line. When a line is required to transmit energy from one point to another, therefore, it is highly desirable, and in fact essential if the utmost efficiency is to be obtained, that the device to which the energy is supplied shall be equal in effective resistance to the surge impedance of the line. Similarly, the internal impedance of the source of energy should be equal to the surge impedance of the line in order that maximum energy shall be transferred to the line by the source. The surge

impedance of a twin wire line is given with sufficient accuracy for practical purposes by the expression :

$$Z_o = 276 \log_{10} \frac{2D}{d}$$

where Z_o = the surge impedance

D = distance apart of centres of wires.

d = diameter of wires.

In consequence of the logarithmic relation the surge impedance is always of the order of 600 ohms for the lines used in practice. For example, if the wires are 0.1 in. in diameter spaced 5 in. apart, $Z_o = 552$ ohms, and if the spacing is increased to 10 inches, or decreased to $2\frac{1}{2}$ inches (the diameter of wire remaining the same), the surge impedance becomes 636 ohms and 444 ohms respectively. For correct termination of a non-resonant line therefore, the impedance of the load should be of this order, while it is also necessary to couple the feeder to the transmitter in such a manner that the effective load imposed by the feeder upon the transmitter is of the same order as the effective resistance of the latter. A radio-frequency transformer with adjustable coupling will ensure that this matching is achieved.

62. A suitable radio-frequency transformer may also be used to couple a non-resonant line to the aerial, but this may not be practicable and is rarely the most convenient method. Fig. 32

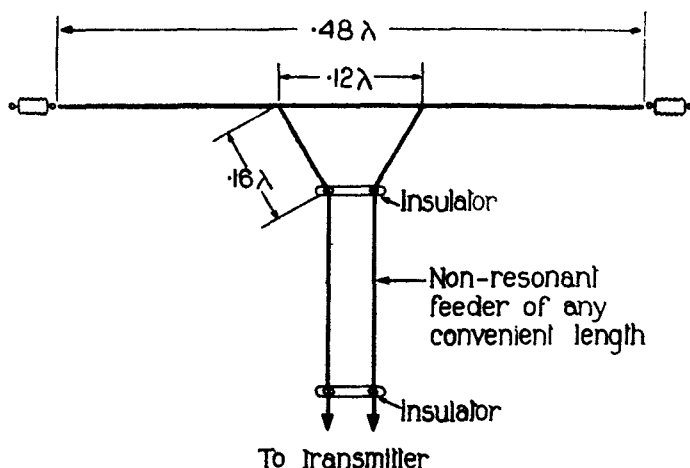


FIG. 32, CHAP. VII.—Auto-inductive coupling between half-wave aerial and non-resonant feeder.

shows a half-wave aerial coupled to a non-resonant line by a kind of auto-inductive coupling. The dimensions given in this figure are appropriate to a 600 ohm line. It is much easier to calculate the spacing necessary to secure this surge impedance than to calculate the tapping points and length of tapered feeder required for a line having a different value of Z_o . A feeder of length $\frac{\lambda}{4}$ may be employed as a radio-frequency transformer between the transmission line and a current-fed dipole. If V_1 and I_1 are the voltage and current at the input end of this "transformation feeder," as it is termed, and V_o and I_o the voltage and current at the output end,

$$V_1 = I_o Z_o$$

$$I_1 = \frac{V_o}{Z_o}$$

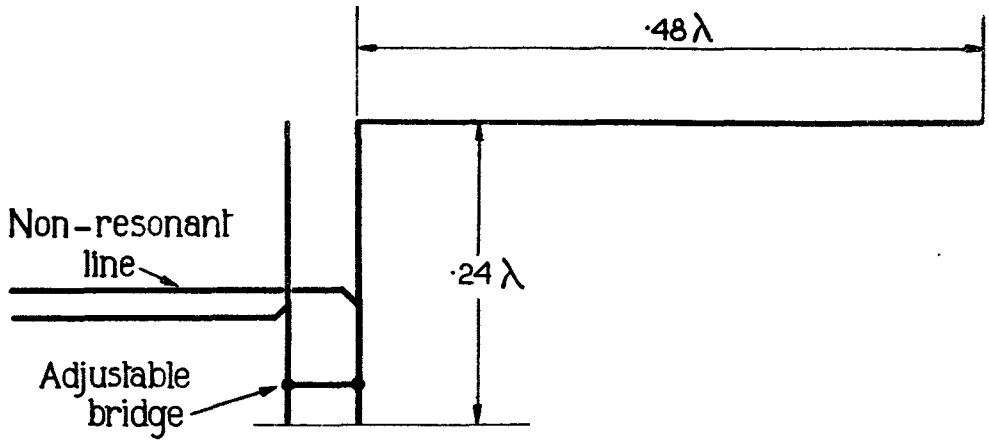


FIG. 33, CHAP. VII.—Voltage fed half-wave aerial with non-resonant feeder.

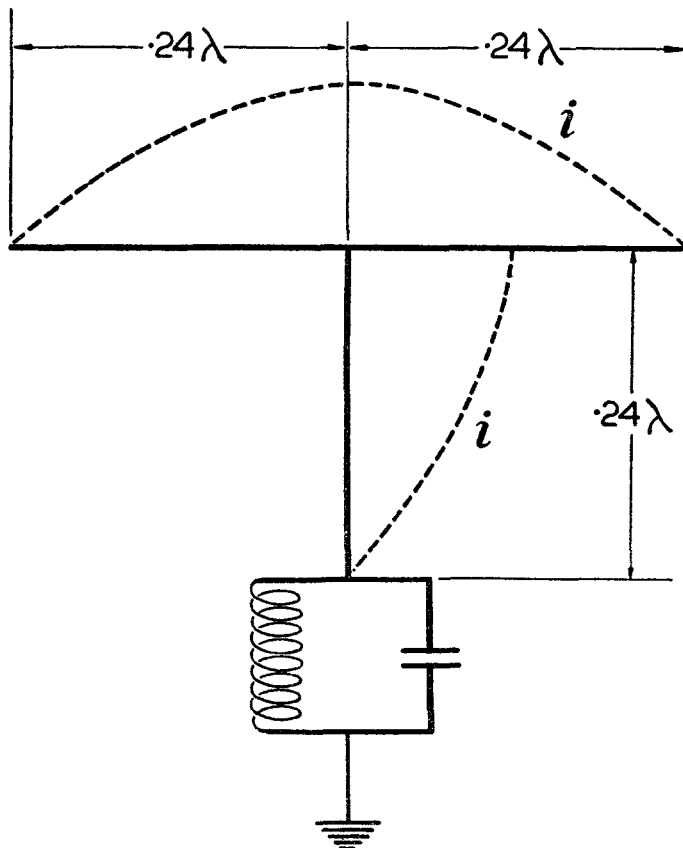


FIG. 34, CHAP. VII.—Basic form of inverted aerial.

Now if R_o is the actual effective resistance of the dipole, $V_o = I_o R_o$. The effective resistance of the line and the load R_o together, measured at the input terminals of the line, will be $\frac{V_1}{I_1}$ or $\frac{I_o}{V_o} Z_o^2$. But $\frac{I_o}{V_o} = \frac{1}{R_o}$, hence the input resistance of the combination, R_i , will be $\frac{Z_o^2}{R_o}$. This signifies that if the surge impedance of the non-resonant line is 600 ohms and the dipole has an effective resistance of 100 ohms, the transformation feeder should have a surge impedance of $\sqrt{R_i R_o}$ or $\sqrt{100 \times 600} = 245$ ohms. This can be achieved by using a feeder composed of four or six wires, i.e. two or three in parallel on each side, the wires being transposed in such a manner that all have uniform capacitance to earth and to every other wire.

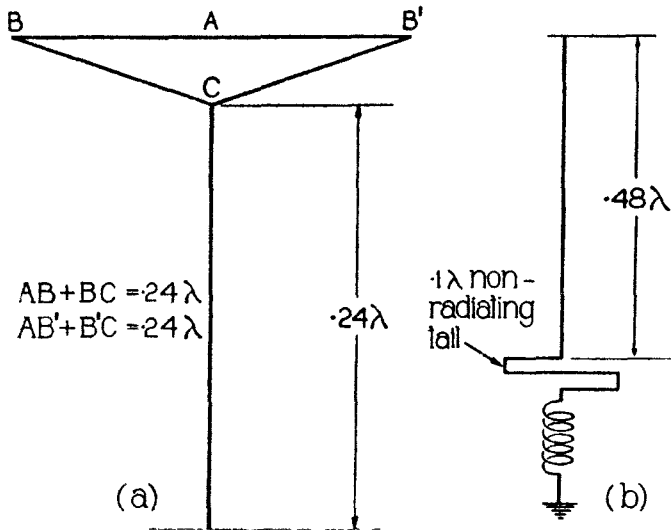


FIG. 35, CHAP. VII.—(a) Development of inverted aerial; (b) Tail feeder.

63. Another convenient method is shown in fig. 33. Here the aerial itself is fed by a resonant $\frac{\lambda}{4}$ feeder, the general arrangement being similar to that of fig. 31. Instead of being directly connected to the transmitter, however, the lower end of the resonant feeder is terminated by a short circuiting bridge, which can be moved from point to point for matching purposes, while the non-resonant line is connected at a point about two-thirds of the length of the resonant line, measured from the upper end. Matching is then carried out by adjustment of the position of the short-circuiting bridge. In order to carry out this adjustment, a low reading thermoammeter (e.g. reading 0-120 milliamperes) is fitted with a rectangular or triangular loop of stiff wire which is secured to a strip of insulating material, the latter being fitted with small hooks so that it may be hung on the feeder wire, the whole assembly forming the secondary winding of a radio-frequency transformer. The absence of stationary waves is indicated by the non-existence of current nodes and loops in the feeder; the current in the line should be practically the same at all points in its length. In the absence of a suitable thermoammeter, a 60 milliamper fuse bulb might be tried, fitting it to the loop in place of the meter. It would perhaps be preferable to remove the screw base and solder the lamp connections direct to the wire loop to avoid the capacitance shunt which would otherwise exist.

CHAPTER VII—PARA. 64

Inverted aeriels

64. In ground to air working over comparatively short ranges it is desirable that the ground station aerial shall radiate most of its energy at a low angle, and it is desirable that the current loops shall be situated as high above ground as possible (cf. fig. 23). At the lower frequencies of the wave-band under consideration, vertical half-wave aeriels are hardly practicable, but a similar vertical polar diagram can be obtained by what is called an inverted aerial, which is defined as an aerial less than $\frac{\lambda}{2}$ in height, but energised in such a manner that it functions as a half-wave aerial. Suppose the wavelength to be 100 metres, an inverted aerial might consist of a vertical wire 24 metres (i.e. $\cdot 24\lambda$) in height with a T or X shaped top each member of which is also 24 metres long (fig. 34). Since the ends of the horizontal members are at the same potential, they may be bent over and interconnected as shown in fig. 35a. The approximate current distribution is shown in fig. 34. The aerial should be fed via a rejector circuit in the same manner as a half-wave aerial. An alternative method of feeding a vertical half-wave or an inverted aerial is to connect a quarter-wave system to its lower end. This consists of a short length (about $\cdot 1\lambda$) of single conductor, which is folded upon itself so as to be non-radiating, which in turn is tuned by means of an inductance at the foot; this coil also serving for coupling to the transmitter. The general arrangement and current distribution are given in fig. 35b.

CHAPTER VIII.—THE THERMIONIC VALVE

ELECTRONIC EMISSION

1. For over a hundred years it has been known that the air in the immediate vicinity of an incandescent conductor also becomes conductive. The effect was studied experimentally by Edison in 1883 and later by Fleming. The apparatus used consisted of an electric lamp, i.e. a glass bulb evacuated in accordance with the best technique of the period, containing a hairpin filament of carbon and a metal plate entirely separate from the filament. The plate could be maintained at any desired potential with reference to the filament, as illustrated in fig. 1. It was found that the space between the filament and the plate became conductive (allowing an electric current to flow) if the plate was maintained at a positive potential with respect to the filament, but was non-conductive if the plate potential was negative with respect to the filament. This unilateral conductivity was later shown to be due to the emission of particles of negative electricity—electrons—from the filament, and their passage to the plate in the presence of a positive potential on the latter.

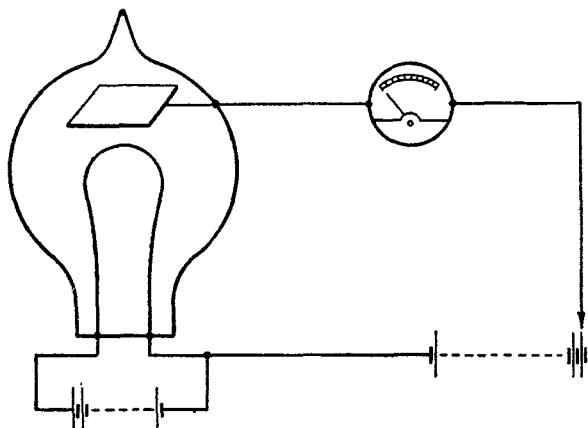


FIG. 1, CHAP. VIII.—The Edison effect.

The mechanism by which electrons are emitted by a hot body was explained by O. W. Richardson in 1901. He assumed that the electrons inside a conductor behave in the same manner as the gas molecules in an enclosed space, that is, they are in a state of continual agitation and random motion, the R.M.S. velocity depending upon the absolute temperature. On the absolute scale, zero temperature is that at which all the particles cease to possess kinetic energy and therefore have no velocity, and is equal to -273°C . The absolute scale of temperature is often referred to as the Kelvin scale, thus $0^{\circ}\text{Centigrade} = 273^{\circ}\text{K}$.

Although the electrons inside a hot body may possess considerable velocity, no electrons leave the boundary surface under normal conditions, because there is at the surface a force tending to hold the electrons within the substance. In order to escape from the atom an electron must do work, and this work must be at the expense of its own kinetic energy. In most instances the amount of work necessary to break the surface tension is greater than the kinetic energy possessed by the electron, so that the latter are bound to the substance. When electrons are detached from a body the phenomenon is known as electronic emission. Electronic emission can be caused in at least three ways:—

- (i) By raising the temperature of the body—*thermionic emission*.
- (ii) By bombardment of a body by electrons or ions—*secondary emission*.
- (iii) By the absorption of electromagnetic radiation of extremely high frequency—*photo-electric emission*.

where ϕ is the equivalent voltage in E.S.U. and e the charge of one electron. To express ϕ in volts this must be multiplied by 300. The two properties required in a cathode material are firstly, low electron affinity, so that a comparatively small temperature rise above the normal will impart sufficient kinetic energy to the electrons to endow them with the ability to break through the cathode surface; and secondly, ability to withstand the necessary rise of temperature without volatilisation or evaporation of material molecules, for this would lead to short life of the cathode as well as irregular behaviour of the valve.

The law of thermionic emission can be shown to be

$$I_s = AT^2 e^{-\frac{b}{T}}$$

where I_s is the emission current in amperes per square centimetre of emitting surface, e is the base of naperian logarithms (≈ 2.71828), and T is the temperature in degrees Kelvin. A is a constant, which is theoretically the same for all pure metal conductors; but in view of the difficulty of obtaining absolutely pure surfaces in practice, the value of A for a number of materials has been determined experimentally by various research workers. The constant b is $1.16 \times 10^4 \times \phi$, ϕ being expressed in volts. Values of A , ϕ and b for different substances are given in the following Table.

	A	ϕ (volts)	b
Tungsten	60	4.53	5.2×10^4
Molybdenum	55	4.58	5.32×10^4
Tantalum	60	4.12	4.77×10^4
Nickel	1,380	5.03	5.84×10^4
Platinum	17,000	6.27	7.27×10^4
Thorium	60	3.35	3.88×10^4
Calcium	60	2.24	2.6×10^4
Magnesium		2.7	3.13×10^4
Mixture of Barium and Strontium Oxides }	.. 0.1 to 0.3	2.3	2.65×10^4
Barium	141	3.58	4.16×10^4
Barium oxide	—	1.12	1.3×10^4
Thoriated Tungsten	3	2.63	3.05×10^4

It is possible that the high values of A for platinum are due to impurities on the surface. It will be observed that for many substances A is about 60, which is one of the values which have been deduced theoretically as the universal constant mentioned above.

The metal tungsten has a melting point of 3540° K. and can be operated at a temperature of 2400° K to 2500° K without fear of volatilisation, giving a copious emission at this temperature as shown in fig. 2. In these respects it is better than any other pure metal.

The anode

4. The action of the second electrode may now be discussed. If this electrode is given a potential positive with respect to the cathode the emitted electrons will be attracted to it, afterwards passing round the external circuit back to the cathode. The plate is then said to function as an anode, and the electron current which is established in the evacuated space and external circuit is called the anode current. The anode may thus be considered to act as a collecting electrode. It is found experimentally that no such current is established if the potential of the plate is negative with respect to the emitting body.

The velocity acquired by an electron in its passage from cathode to anode depends upon the magnitude of the potential difference between the electrodes, or more accurately upon the potential gradient. The potential gradient is measured in volts per centimetre and is synonymous with the electric field strength. This velocity may be estimated by the application of fundamental principles, as follows.

CHAPTER VIII.—PARA. 5

The force F exerted by an electrostatic field upon a charged body is equal to the product of the charge q carried by the body and to the electric field strength Γ or

$$F = q\Gamma \times 10^7$$

where F is the force in dynes, q the charge in coulombs and Γ the electric field strength in volts per centimetre. This force acts in the direction of the field at the point occupied by the charge and causes an acceleration in this direction. Provided that the velocity of the body does not approach that of light the ordinary laws of mechanic may be applied to find the acceleration and since $F = ma$ (see Introduction)

$$a = \frac{\text{force in dynes}}{\text{mass of body in grams}}$$

a being then given in centimetres per second per second. The velocity of the moving body will increase or decrease during its passage through the field, according to the relative signs of field and charge. The kinetic energy acquired is equal to the product of the field strength Γ , the distance moved through, d , and the magnitude of the charge, q . If d is in centimetres and q in coulombs, the kinetic energy is W ergs, where

$$\begin{aligned} W &= dq\Gamma \times 10^7 \\ &= Vq \times 10^7 \end{aligned}$$

because $\Gamma d = V$, the P.D. between the points of arrival and departure of the charge. The latter is said to fall through this potential difference.

Suppose the charge to start from rest under an attractive force, attaining a final velocity of u centimetres per second. The kinetic energy gained during its motion is $\frac{1}{2} mu^2$ ergs, hence

$$Vq \times 10^7 = \frac{1}{2} mu^2$$

$$\text{and } u = \sqrt{\frac{2 \times 10^7 Vq}{m}} \text{ cm. per second}$$

Example 1.—The charge on an electron is 1.59×10^{-19} coulomb and its mass 9×10^{-28} gram. Calculate the velocity gained by an electron in falling through a P.D. of 10 volts.

$$\begin{aligned} u &= \sqrt{\frac{2 \times 10^7 \times 10 \times 1.59 \times 10^{-19}}{9 \times 10^{-28}}} \text{ cm. per second} \\ &= 1170 \text{ miles per second.} \end{aligned}$$

Owing to the high velocity attained by electrons when moving through space comparatively free from matter, it is convenient to express the velocity in terms of the voltage through which the electron has fallen. It is obvious that a body of greater mass, such as a hydrogen ion, will acquire a much lower velocity for the same P.D. In the above example, substitution of the mass

of a hydrogen ion will give $\frac{1170}{\sqrt{1840}}$ or 27 miles per second.

5. The assembly so far considered, consisting of an evacuated space enclosed by a container called the envelope, a cathode or emitting electrode and an anode or collecting electrode, is called a two-electrode valve or diode. Its essential property is that of unilateral conductivity.

The original function of the diode was that of a detector of radio-frequency currents in a wireless receiving aerial, and is dealt with in Chapter X. It was later found that by the addition of other electrodes the valve could be rendered much more sensitive and endowed with amplifying properties. The valves so evolved are often designated by the total number of active electrodes they possess, e.g. a valve possessing three active electrodes is called a triode, four active electrodes a tetrode, and so on. Two or more independent electrode assemblies are sometimes enclosed in a single envelope, for example a diode-triode possesses a single cathode, an anode which in conjunction with the cathode constitutes a diode, and an additional anode and grid forming with the cathode a triode which functions quite independently of the diode assembly.

The following paragraphs, dealing briefly with the design of the cathode and the general construction of valves, are applicable to all types in common use. The cathode is invariably heated by electrical means although this is merely a matter of practical convenience. Two methods of achieving the desired end are in common use, the resulting designs being known as "directly heated" and "indirectly heated" cathodes respectively.

The directly heated cathode

6. This consists of a filament of wire, which is heated by passing through it a current of electricity obtained from dry cells, accumulators, or in certain instances from the supply mains. This filament current plays no part whatever in the action of the valve, other than providing the necessary heating. The material of the filament is generally either pure tungsten, thoriated tungsten or oxide coated wire.

Pure tungsten has an operating temperature 2400° K to 2500° K. It is drawn into filaments of from .05 mm. diameter for use in small receiving valves, to 1.3 mm. in large transmitting valves, the corresponding heating currents being from .5 to 75 amperes. Pure tungsten filaments are not used in service receiving valves.

Thoriated tungsten.—Thorium was originally added to tungsten lamp filaments in order to render them less brittle, and it was discovered that these filaments gave a copious emission of electrons at a lower temperature than pure tungsten. The actual substance added is thoria (thorium oxide), which forms a "solid solution" in the tungsten. When the filament is heated, the oxide is reduced, and molecules of pure thorium are deposited on the surface of the filament, forming a kind of tube which is rich in electrons of a low affinity. Hence such filaments give ample emission when heated to a temperature of from 1800° K to 1900° K. This thin layer of thorium is easily destroyed by excessive temperature or by bombardment by positive ions formed from gas molecules in the envelope. The filament then fails to emit at its normal temperature.

If a thoriated-tungsten filament loses its emission, it may be "reactivated" by applying two and a half times the normal filament voltage for 20 seconds with no anode potential. This will decompose some thoria and so provide a supply of thorium molecules. If a filament voltage of about 20 per cent. above normal and an anode voltage not exceeding 20 volts is now applied for a short period the thorium molecules will be distilled out to the surface of the filament. The thinnest thoriated tungsten filament is designed to emit with a filament current of .06 amp.

Some service transmitting valves are made with thoriated tungsten filaments although they function as "bright emitters." The admixture of thoria in this case is to reduce the brittleness of the filament rather than to increase the emission. Others such as the V.T.25 valve, have thoriated tungsten filaments designed to run at a dull red heat. These valves are given a hydrocarbon treatment, being run in an atmosphere of volatilised naphthalene or some similar compound during a certain stage in the manufacture. This converts some of the surface tungsten into tungsten carbide, and the filament will then stand up to considerable positive ion bombardment without disintegration of the emitting surface, but the process tends to make the filament more brittle than the ordinary thoriated tungsten filament and this should be borne in mind when handling these valves.

Oxide coated filaments.—The oxides of barium, strontium and calcium will emit copious supplies of electrons at a dull red heat, about 1200° K. These substances cannot be drawn into wire but can be deposited upon a core of some metal, for example, platinum with 10 per cent. iridium or 5 per cent. nickel, commercial nickel, or an alloy of nickel, cobalt, iron and titanium called Konel metal. The coating material frequently used is a mixture of equal weights of barium carbonate and strontium carbonate, ground to a fine powder and thoroughly mixed, being then formed into a thin paste or paint by the addition of water or amyl acetate. The coating is applied by drawing the wire through a series of baths containing the oxide, a drying process occurring between each bath. The latter process consists of baking for a few seconds at a temperature of about 700° C. in an atmosphere of carbon dioxide. The correct thickness of

CHAPTER VIII.—PARAS. 7-9

coating is maintained by weighing samples of the coated materials, and when sufficiently coated the surface is given a protective coating of paraffin wax or of collodion. This type of emitter requires activation after being fitted in the valve, in order to reduce the carbonates to oxides, as described in the section on valve manufacture.

The indirectly heated cathode

7. If alternating current is used to heat a directly heated cathode, the variations of voltage between its ends produce fluctuations of the anode current at the supply frequency, producing a hum in the telephone receivers or loud speaker. This effect is overcome in the indirectly heated type by making the cathode and heating device electrically independent, the cathode being heated by radiation and conduction of heat from the "heater." The cathode is an oxide-coated metal tube or thimble, while the heater takes the form of a spiral or hairpin filament (fig. 3),

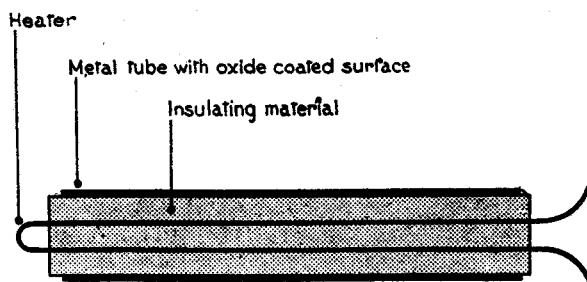


FIG. 3, CHAP. VIII.—Indirectly heated cathode.

which is connected to the A.C. supply. The filament is kept in position in the thimble, and insulated from it, by a packing of porcelain, magnesia or silica. The heater is not regarded as one of the active electrodes of a valve, thus an indirectly heated diode has two active electrodes, viz. anode and cathode, and also a heater for the latter, four external connections being therefore necessary.

Anode dissipation

8. When electrons emitted by the cathode impinge upon the anode, their kinetic energy is almost entirely converted into heat, causing an increase in the temperature of the anode. Suppose that n electrons of charge e coulombs reach the anode every second, having fallen through a P.D. of V volts. The rate of conversion of kinetic energy into other forms is neV joules per second. For example, if $n = 10^{18}$ and $V = 1600$ volts, since $e = 1.59 \times 10^{-19}$ coulomb, the energy converted per second is

$$\begin{aligned} P &= 10^{18} \times 1.59 \times 10^{-19} \times 1600 \\ &= 254 \text{ watts} \end{aligned}$$

In practice the anode current is measured in milliamperes, e.g. 159 milliamperes in the above example. The power expended in the form of heat is then equal to the product of the anode current I_a and the anode-filament P.D. V_a or

$$P = \frac{I_a V_a}{1000}$$

The rating of a power valve is the amount of power which the valve can safely dissipate in the form of heat. The power supplied to the valve will generally—always in practical working circuits—be greater than this, because some of the electrical energy supplied to the valve will be transformed into some other form of useful energy and not into heat.

9. The anode, and any additional electrodes which the valve may possess, must be constructed from a material which will withstand the heat generated in manufacture and use, and

will not retain gas molecules in adsorption. As the heat evolved can only be lost by radiation to the walls of the envelope, the anode must possess good radiating properties and must also be

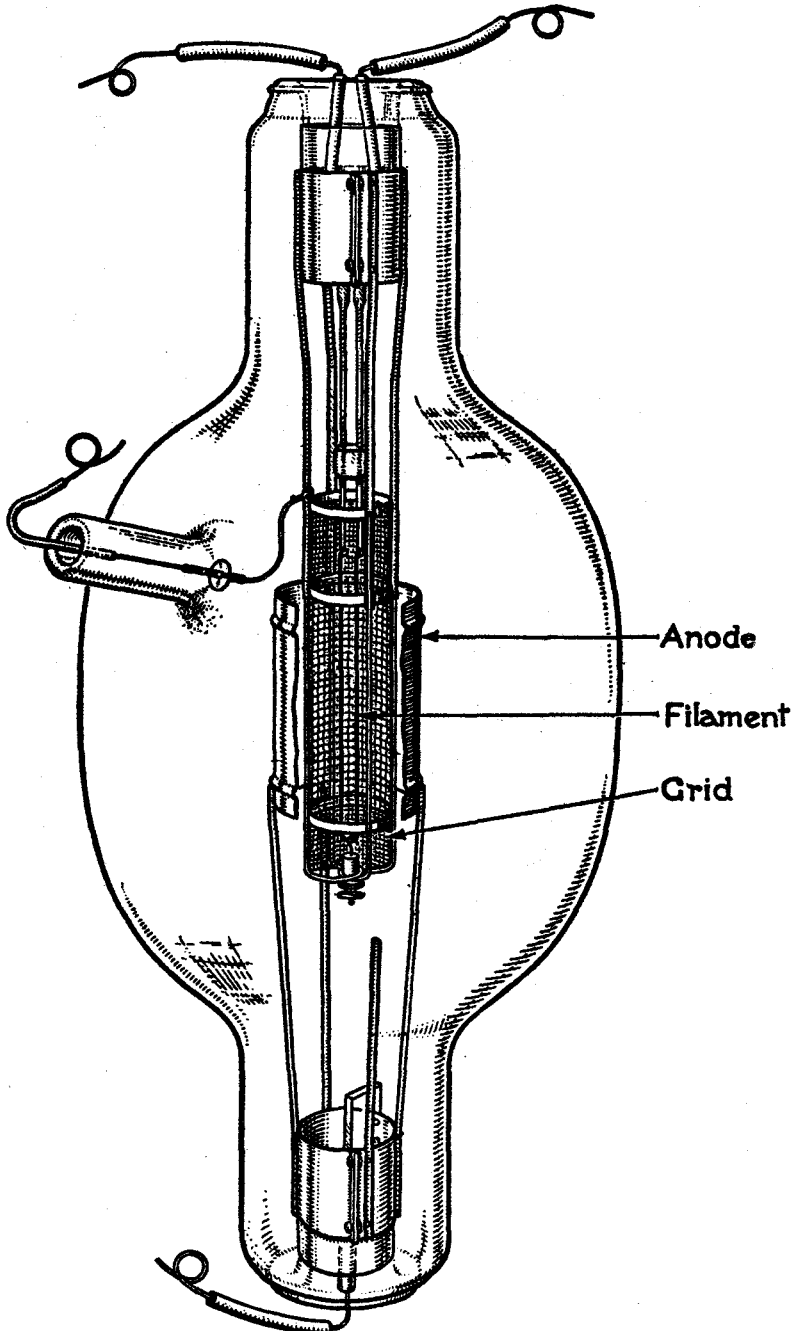


FIG. 4, CHAP. VIII.—Typical transmitting valve.

CHAPTER VIII.—PARAS. 10-11

capable of taking and maintaining the desired shape and size. It is essential that no appreciable amount of gas is liberated by the anode when operating at its normal temperature, for if this is the case, gas molecules will become ionised by collision with electrons, and the result is an increase of anode current, with further rise of anode temperature and increased liberation of gas molecules. The effect is therefore cumulative and may lead to the establishment of an arc discharge between anode and cathode. The nature of the gas is of importance in valves having filaments of thoriated tungsten. The electronic emission is due to a very thin layer of thorium on the surface of the filament, and a very small amount of oxygen is sufficient to oxidise this layer which then loses its emission property to a considerable extent.

For receiving valves and low power transmitting valves, nickel is largely used for the construction of anodes and other electrodes. For higher powers, the metal molybdenum is largely employed. If heated to about 1400°C . during the evacuation process, the adsorbed gas is largely given off, and the anode may be operated at a working temperature as high as 1000°C . without further liberation of gas. The mechanical properties of molybdenum are not ideal, and it is difficult to obtain in large thin sheets such as are used to form nickel anodes. In certain types of valve the anode is formed of molybdenum strip about 2 millimetres wide, which is woven into a basket-work structure, and is more easily degassed during evacuation than a solid electrode. It also warps to a much smaller extent when heated, and tends to give constancy of inter-electrode capacitance which is of considerable importance in operation.

During the last few years, graphitic carbon has come into use to some extent. The advantages claimed are (i) its high mechanical rigidity and (ii) good heat-radiating properties. Against this must be set a greater tendency to gas adsorption, although it is claimed that this difficulty can be overcome by suitable treatment during manufacture.

10. The envelope is usually of glass, at any rate up to a rating of 1000 watts (e.g. V.T.9B). A glass valve typical of the largest size made is shown in fig. 4. Larger envelopes are made of silica (fused quartz). This material has a negligible coefficient of expansion with temperature, and does not soften unless heated to about 1400°C . This leads to a smaller valve for a given power, since the envelope may be brought nearer to the heated anode without ill effects.

The silica valve can be cut open with a carborundum wheel for replacement of a burnt-out filament, after which it is re-assembled and evacuated. Defective silica valves are thus of definite value, and should be treated as carefully as new ones.

The "cooled anode" valve—cooled anode transmitting or C.A.T. and cooled anode modulating or C.A.M.—has for its anode a copper tube which forms part of the envelope and may be cooled by water or air circulation. The success of this valve is entirely due to the development of a method of making a gas-tight seal between glass and copper.

Evacuation

11. The evacuation process aims at the elimination of all occluded gas, as well as that filling the envelope itself. All the internal parts must be chemically clean after assembly, when the valve is ready for pumping. This process is carried out individually for large valves, but small receiving valves are pumped by semi-automatic "mass-production" processes. With this qualification, the general features of the evacuation process are the same for all valves. The pumping outfit consists of a series of pumps, the general arrangement being shown in Fig. 5. The first, or rough pump, consists of an air pump of ordinary piston type, and is capable of reducing the pressure to about $\cdot 01$ mm. of mercury. This rough pump is followed by a rotary pump, which in turn reduces the pressure to some $\cdot 0001$ mm. The rotary pump in turn is followed by a "mercury vapour" pump, the action of which is as follows. In a metal flask is a pool of mercury, which is vaporised by a bunsen flame. The molecules of mercury vapour rush up the flask, and on passing the mouth of a tube to which the valve is connected create a partial vacuum, and thus gas molecules are drawn from the valve into the flask. The mercury vapour condenses on the cold upper part of the flask, and trickles back to the pool to undergo further vaporisation. A liquid air trap is inserted between the valve and the mercury pump in order to "freeze out" any few molecules of mercury vapour which might find their way back towards the valve.

The final pressure in the valve may be less than $\cdot 0000001$ mm. of mercury. During the pumping process, the valve is raised to as high a temperature as possible consistent with the envelope remaining rigid, in order to remove occluded gas from the envelope. In the later stages of evacuation, the metal parts inside the valve are heated to about 400° C. either by "electronic bombardment" or by eddy currents induced by means of radio frequency current in a small coil surrounding the valve, with the same object.

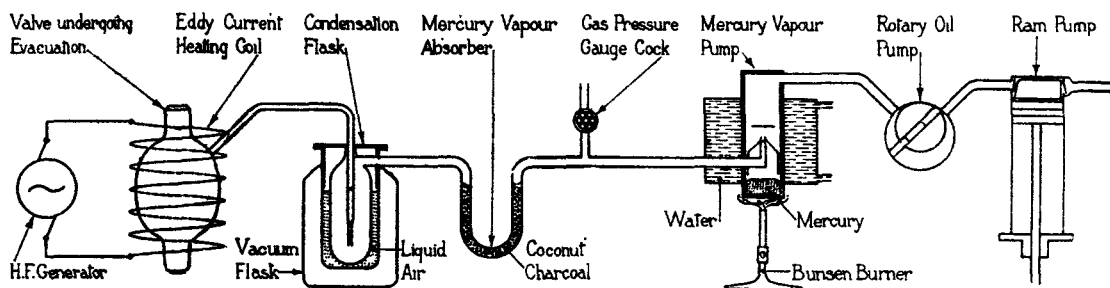


FIG. 5, CHAP. VIII.—Evacuation of transmitting valve.

Activation.—If the valve is fitted with an oxide-coated emitter, the latter must now be activated. A heating power of about double the normal, i.e. that at which the heater will be run in actual service, is applied to the cathode for several minutes, then somewhat reduced and maintained until evolution of gas from the cathode becomes negligible. The pumping process is kept up during this period and when the gas pressure has been reduced to about 10^{-5} mm. of mercury, the activation is usually complete.

Gettering.—The gas which is most detrimental to the proper functioning of the valve is oxygen, and the residual traces of this gas are eliminated by causing it to combine with a small quantity of magnesium, calcium, or barium which is placed inside the bulb. When the temperature of the bulb and electrodes is raised to a certain degree by the eddy current treatment described above, the metal combines with the residual oxygen and the resulting compound is deposited upon the inside of the bulb, causing the silvery or bronzed appearance characteristic of a "dull-emitter" valve. This process is applied to both thoriated-tungsten and oxide-coated filaments but not to transmitting valves, owing to the high temperature at which they operate which would cause re-volatilisation of the deposit and consequent lowering of the vacuum.

Metallising.—A deposit of soft metal is frequently sputtered over the outside of the envelope of receiving valves to assist or complete the screening of circuits by confining oscillatory electric or magnetic fields proper to the valve to its interior and by excluding unwanted fields from the vicinity of the electrode system. The metallising is usually connected to one filament pin, i.e. that one which is at or near earth potential, or in some instances to a separate pin.

Hard, gas-filled and soft valves

12. In the foregoing outline of valve manufacture, it has been assumed that the finished valves are exhausted to the highest degree of vacuum attainable. This may be said of all valves in general use for transmission and reception in which the presence of gas is undesirable, and they are known as "hard" valves. In some valves gas molecules are introduced deliberately, after the exhaustion has been carried out and these are known as "gas-filled" valves. Examples are the "mercury vapour diode" and the "thyatron." When a gas-filled valve is made, it is first thoroughly exhausted, and a definite quantity of the desired gas afterwards introduced.

A hard valve may become "soft" in use, i.e. the degree of vacuum lower than normal owing to some imperfection in manufacture or irregular usage in its subsequent life. The signs of this are "hot spots" on the filament and possibly a "blue glow" in the valve when its anode

CHAPTER VIII.—PARAS. 13-14

potential is raised to nearly its working value. This blue glow is due to ionisation of the gas molecules by the impact of the travelling electrons. A valve which shows such signs of softness should be replaced as soon as possible, otherwise it may cause a wireless failure at a critical moment. The blue glow due to ionisation forms an aurora round the cathode, which may extend and fill the space between cathode and anode; a slight bluish fluorescence may exist in localised places on the anode of an indirectly heated valve, but this is not harmful.

13. When the diode is connected as shown in fig. 1, electrons flow through the exhausted space from filament to anode. The anode current (I_a) thus established is, in general, less than that calculated from the emission equation (para. 3 and fig. 2). Although the emission given by this formula actually occurs, not all the emitted electrons reach the anode, a large number returning to the cathode in spite of the attraction of the anode. The cause of their return is the "space charge," which consists of the cloud of emitted electrons occupying the interelectrode space. This is in effect an electric field of negative sign, and therefore tends to repel the electrons which are leaving the cathode, causing them to return to it. The anode current of the valve is thus influenced both by the positive potential of the plate and the negative potential of the space charge. In fig. 6 an attempt is made to show the distribution of the electric field between the

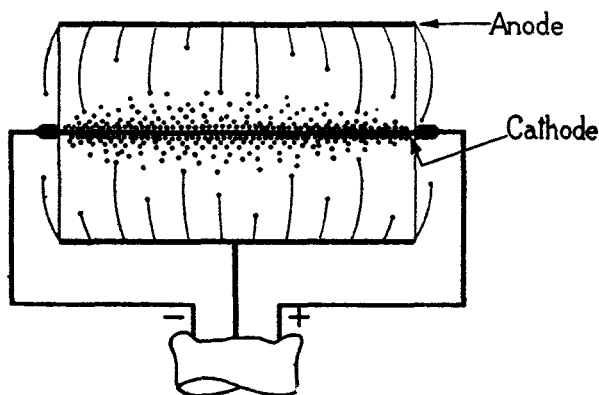


FIG. 6, CHAP. VIII.—Space charge in diode.

cathode and anode when the cathode is emitting freely and the anode is maintained at a positive potential with respect to it. In some instances the presence of gas causes a positive space charge, because the emitted electrons colliding with gas molecules, dislodge electrons from the latter. These electrons then join the stream flowing to the anode, while the positive ions (gas molecules) are attracted to the filament, impinging on it with considerable velocity, and causing the "hot spots" mentioned above. It will be seen that a soft valve will, with similar operating potentials, give a larger anode current than a hard valve of identical design.

Characteristic curve of diode

14. If the cathode of a two-electrode valve is maintained at its normal temperature, i.e. that which will provide ample emission, and the P.D. between anode and cathode is varied in steps, the relation between "anode-cathode P.D." and "anode current" can be exhibited in graphical form as shown in fig. 7. This graph resembles the B/H curve shown in fig. 16, Chap. II, and its shape can be explained in similar terms. Thus for low values of anode voltage, the resulting attraction of the emitted electrons is nearly overcome by the repulsive effect of the space charge, and the anode current is correspondingly small. When the anode voltage has been raised to such a value that the space charge is entirely annulled, an increase of anode voltage leads to a proportional increase in anode current. When the anode voltage is raised above a certain value, however, the anode current no longer increases proportionally, because nearly all

the emitted electrons are already reaching the anode, and an increase of anode voltage cannot increase the number of electrons emitted. The current is then said to have reached "saturation value" or "full emission." Oxide-coated filaments do not give a well marked saturation value, nor, to a less extent, do thoriated tungsten filaments.

The relation between anode current (I_a) and anode voltage (V_a) in the curve of fig. 7, is given by the equation

$$I_a = AV_a^{\frac{3}{2}}$$

where A is a constant depending upon the design of the electrode system. This law is only true provided that the emission is sufficient to avoid saturation, i.e. when the anode current is limited by the space charge. As an example of the practical use of this formula, consider the design of a pure tungsten filament for a large diode with cylindrical anode. For pure tungsten the constant A is

$$\frac{2.92 l}{\beta^2 d \times 10^3}$$

where l is the length (centimetres) of the filament.

d the diameter (centimetres) of the anode.

β a function of the ratio $\frac{\text{anode radius}}{\text{cathode radius}} = R$ which is approximately unity for all values of R greater than 10.

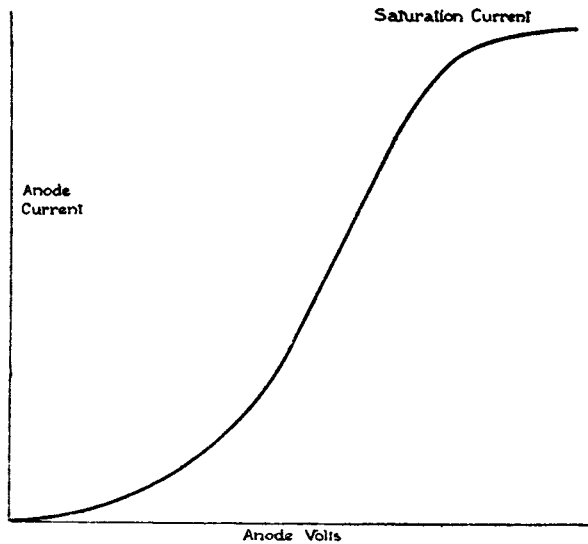


FIG. 7, CHAP. VIII.—Characteristic curve of diode.

Example 2.—A tungsten filament is required to give an anode current of 500 milliamperes at 400 volts, the anode diameter being 2 cm. and the filament diameter 1 mm. Find the length of filament necessary.

Since $R > 10$, $\beta^2 \doteq 1$.

$$\begin{aligned} I_a &= \frac{2.92 \times 400^{\frac{3}{2}} \times l}{2 \times 100} \\ l &= \frac{2 \times 100 \times 500}{2.92 \times 8000} \\ &= 4.3 \text{ cm.} \end{aligned}$$

CHAPTER VIII.—PARA. 15

In the service the diode is used.

(i) As a power rectifier, for supplying high voltage of steady value from alternating supply mains. Both the "hard" diode and the "mercury vapour" diode are used for this purpose.

(ii) As a "limiting" device in certain Transmitter-Receivers.

Both these functions will be described in the appropriate chapters.

THE TRIODE

Introduction of the grid

15. It has been observed that if the cathode of a valve is maintained at a temperature giving ample emission, and the anode is maintained at some potential positive to the cathode, the electron current is limited by the negative space charge, while if owing to the presence of gas molecules a positive space charge is formed, the anode current is increased. This at once suggests that a control of the anode current could be obtained by means of a space charge of variable

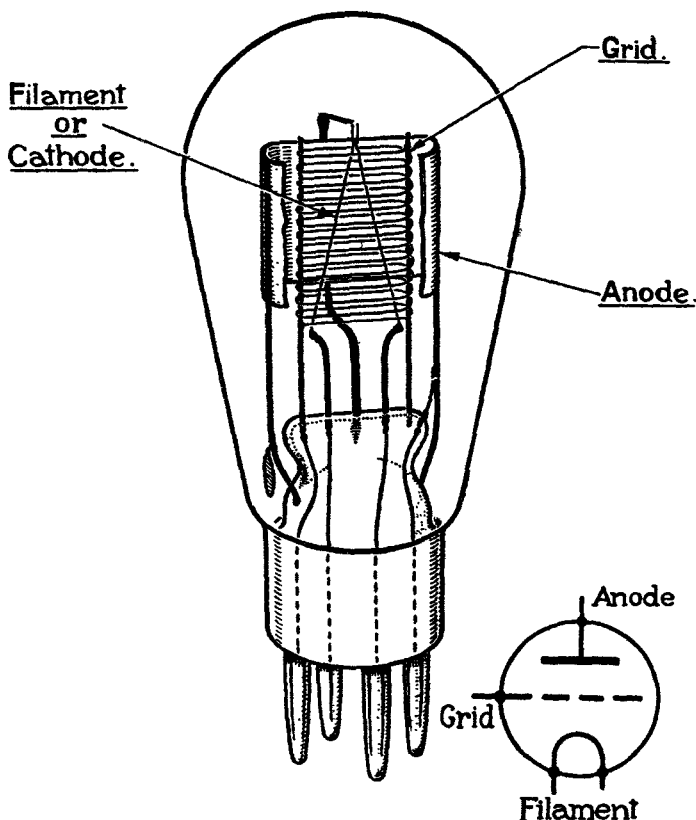


FIG. 8, CHAP. VIII.—Construction of triode.

amount and sign. The equivalent of this can be achieved by inserting a third electrode, whose potential is variable, between cathode and anode. This electrode must be perforated in order that the electron flow is not entirely obstructed and in practice is in the form of a grid or wire spiral. The component parts of a directly heated three-electrode valve or triode are shown in fig. 8, while fig. 9 shows two standard holders used for these valves and is self-explanatory.

Different specifications are necessary for triodes which perform various functions, but these fall into five main classes :—

- (i) Receiving triodes for general amplification.
- (ii) Receiving triodes specially selected as detectors.
- (iii) Receiving triodes for power amplification.
- (iv) Transmitting triodes for low power (e.g. aircraft) transmitters.
- (v) Transmitting triodes for high power (e.g. ground station) transmitters.

The following general theory is the same for all these unless specifically stated to the contrary.

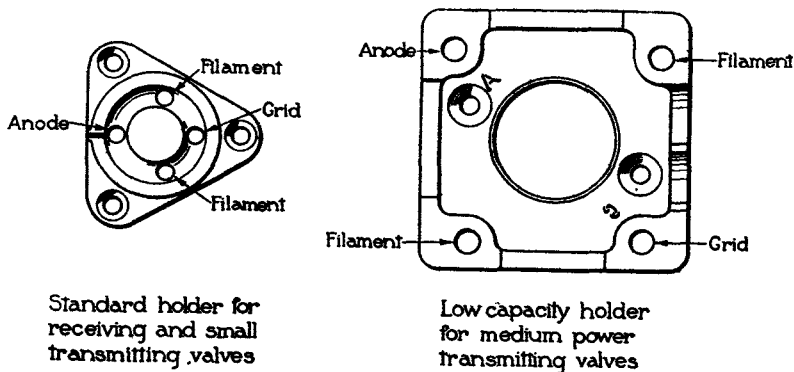


FIG. 9, CHAP. VIII.—Valve holders.

Characteristic curves of the triode

16. A graph which exhibits the relationship between the current flowing to any electrode, and the P.D. between any one electrode and the cathode, is called a static characteristic curve. It is necessary to specify that any electrode whose potential variation is not the subject of investigation shall be maintained at some constant potential, during the process of obtaining the points for plotting one curve.

Fig. 7 is actually the “anode current—anode voltage characteristic” of a diode. A triode has four such characteristics, viz :—

- (i) The Anode Current—Anode Volts Curve ($I_a - V_a$) (the grid voltage being maintained constant).
- (ii) The Anode Current—Grid Volts Curve ($I_a - V_g$) (the anode voltage being maintained constant).
- (iii) The Grid Current—Grid Volts Curve ($I_g - V_g$) (the anode voltage being maintained constant).
- (iv) The Grid Current—Anode Volts Curve ($I_g - V_a$) (the grid voltage being maintained constant).

It must be noted, however, that there is an indefinite number of possible curves in each “family” denoted by (i), (ii), (iii), (iv) above, each curve corresponding to one particular value of the potential of the electrode whose voltage is fixed. The latter is called a “constant parameter” for each curve of a family. Characteristic curves are used to explain the action of the valve under given conditions, and to determine suitable operating conditions for any desired purpose.

17. The characteristic curves of a given valve are obtained by plotting the current and voltage relations, which are observed by means of the circuit shown in fig. 10, in which it will be seen that it is possible to apply variable voltages between grid and cathode and between anode and cathode, as well as to vary the emission of the latter by variation of the current through the

CHAPTER VIII.—PARAS. 18-19

filament. By an established convention, all P.D.s. are specified with reference to the cathode, or to the negative end of the filament when a directly heated cathode is used. It is important to connect the milliammeter I_a and the microammeter I_g in the positions shown, so that they measure only the actual current flowing at the anode and grid respectively.

The change-over switch (S) is fitted in order that the grid potential may be made either positive or negative to the filament. It is not usual to provide a similar switch for the anode potential, but it should be verified by reversing the necessary connections, that if the anode is given any potential negative to the filament, no anode current will flow.

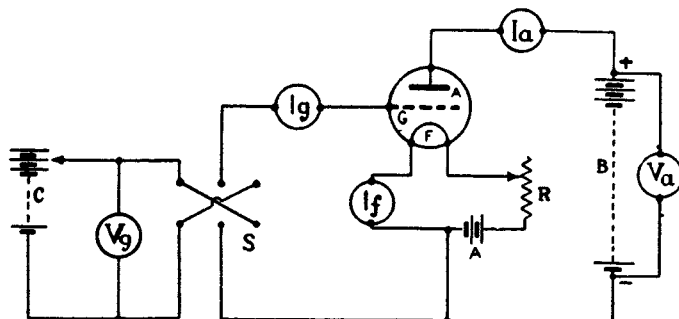


FIG. 10, CHAP. VIII.—Circuit used to obtain triode characteristics.

The various power supply devices used with all valve circuits may conveniently be defined with reference to this figure. The battery (A), supplying the heating current for the cathode, is called the L.T. (low tension) battery. The battery (B) applying a P.D. between anode and cathode, is called the H.T. (high tension) battery, and the battery (C) the grid bias battery. In American literature they are designated the A, B and C batteries respectively.

"Constants" of the triode

18. From the curves obtained by the use of the valve characteristic circuit it is possible and usual to derive numerical data which serve as a basis for the design of appropriate circuits for use with any particular valve, and to compare the merits of different valves for any special purpose. These data are usually called the constants of the valve, but it must be clearly understood that they are not constant over a wide operating range. These constants are:—

(i) The anode A.C. resistance, symbol r_a , which is the ratio of a small change of anode voltage to the corresponding change of anode current as determined from the static $I_a - V_a$ characteristic, the grid voltage and electron emission remaining constant.

The anode A.C. conductance, symbol g_a is the reciprocal of this: $g_a = \frac{1}{r_a}$.

(ii) The mutual A.C. conductance, symbol g_m which is the ratio of a small change of anode current to the corresponding change of grid voltage as determined from the static $I_a - V_g$ characteristic, the grid voltage and electron emission remaining constant.

(iii) The amplification factor, symbol μ , which is the numerical ratio of the slope of the $I_a - V_g$ curve to the slope of the $I_a - V_a$ curve, the slope in each case being taken at the point representing the particular adjustment under consideration.

$$\text{Algebraically, } \mu = \frac{g_m}{g_a} = g_m r_a$$

The $I_a - V_a$ characteristics. Derivation of r_a and g_a

19. A family of $I_a - V_a$ characteristics for a typical receiving valve having an oxide-coated filament is given in fig. 11, curves being drawn for various values of V_g , viz:— zero, +2 volts, -2 volts and -4 volts respectively. It will be observed that the curves are approximately straight and parallel over a wide range, the lower limit being in the neighbourhood of two

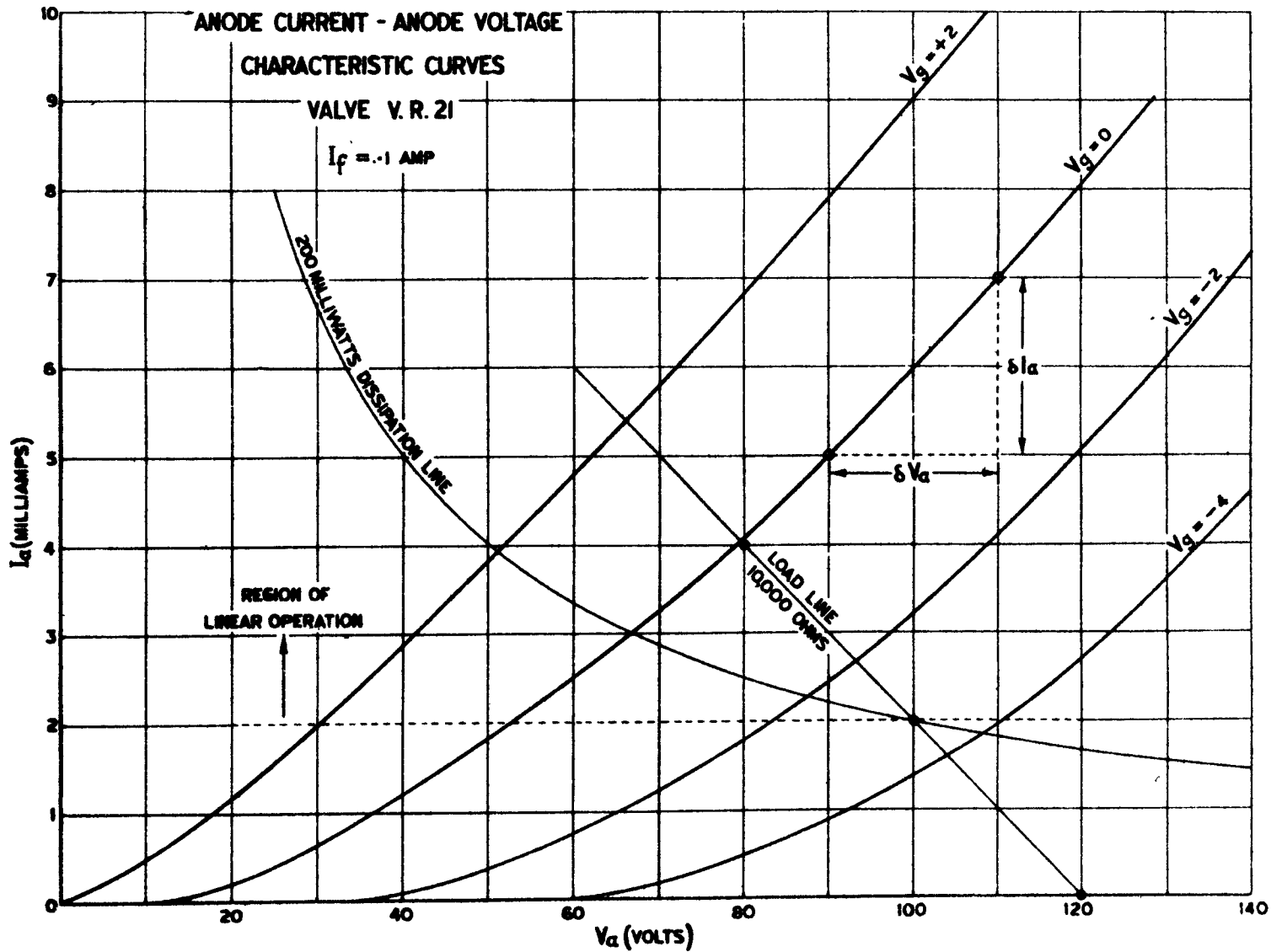


FIG. II
CHAP. VIII

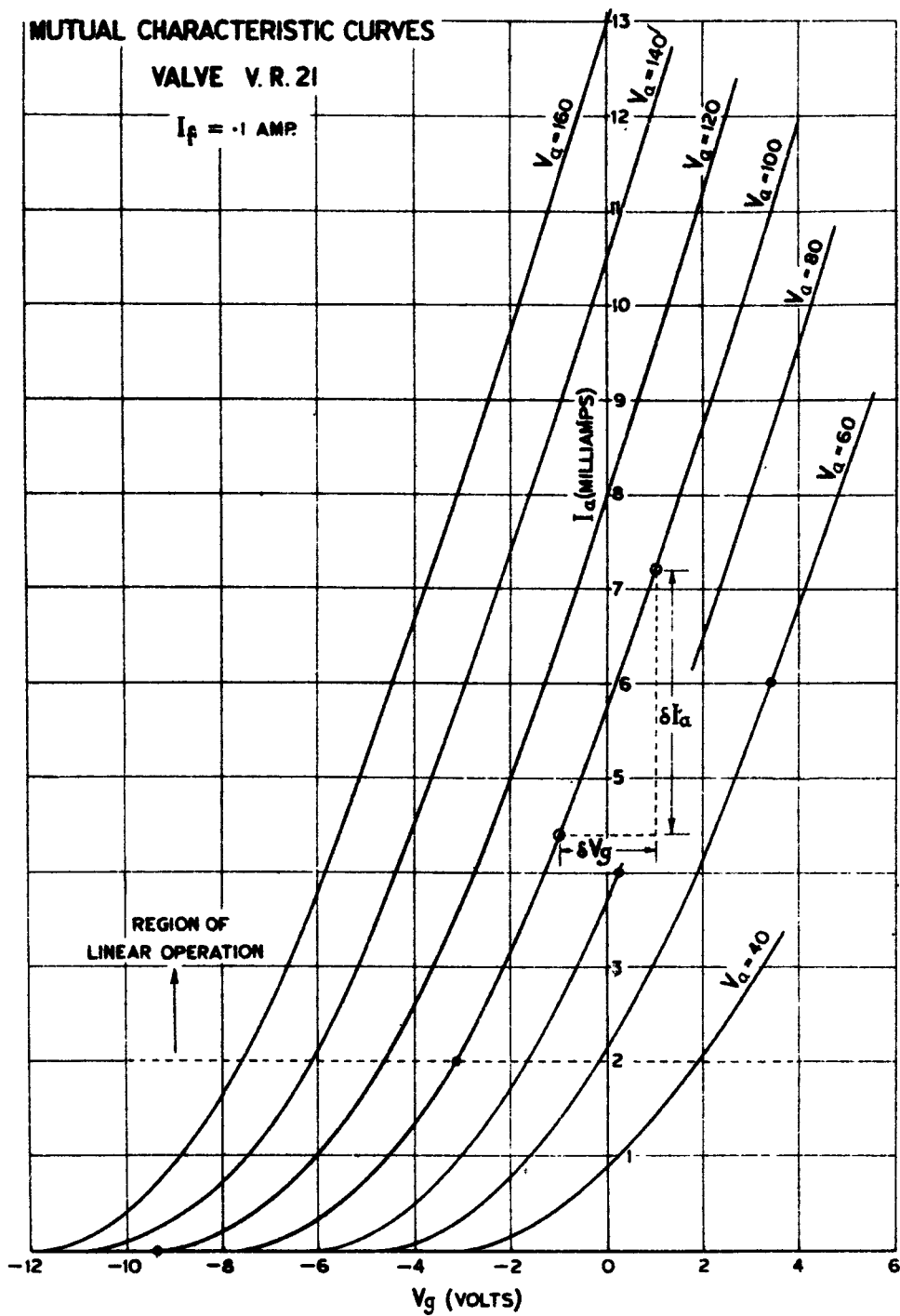


FIG. 12
 CHAP VIII

milliamperes of anode current. The upper limit is not shown, because with this type of cathode there is danger of exceeding its safe emission if higher operating potentials than those shown are applied. This region of approximate straightness and parallelism is termed the region of linear operation. The effect of increasing the negative potential applied to grid is to shift the curve bodily to the right, while slightly decreasing its slope.

The anode A.C. resistance can be obtained from any curve in this family, depending upon the intended operating conditions. In order to illustrate the procedure, the operating potentials have been taken as $V_g = 0$, $V_a = 100$. The corresponding mean value of I_a is 6 milliamperes. Bearing in mind the definition of r_a , mark off on the graph a small change of V_a disposed symmetrically about the mean operating point, calling this small change δV_a . The corresponding change of I_a , called δI_a , is then ascertained, and the value of r_a follows from the definition, being the ratio $\delta V_a / \delta I_a$. In the figure $\delta V_a = 20$ volts, $\delta I_a = 2$ milliamperes, therefore $r_a = 20$ volts $\div \frac{2}{1000}$ amperes = 10,000 ohms.

The anode A.C. conductance is the reciprocal of this, i.e. $\frac{1}{10,000}$ siemens or $\cdot 1$ milli-mho.

The $I_a - V_g$ characteristic. Derivation of g_m

20. A typical family of these curves is shown in fig. 12, being the mutual characteristics of the valve previously discussed. A separate curve is shown for each step of 20 volts anode potential in the range 160–40 volts, these being the limits between which the valve can be usefully operated. The region of linear operation should be observed, and also that the curves for 60 and 40 volts tend to a decreasing slope, owing to the fact that when the anode potential is low and the grid potential is positive, the grid becomes a collecting as well as a controlling electrode.

The mutual A.C. conductance can be obtained from any one curve of this family; the chosen curve should of course agree with the operating conditions previously prescribed. The procedure is shown in the figure. A small change of V_g , (δV_g), having been marked off, the corresponding change of I_a , (δI_a) is measured. By definition, $g_m = \delta I_a / \delta V_g$. In the example, $\delta I_a = 2\cdot7$ milliamperes, $\delta V_g = 2$ volts, hence $g_m = 2\cdot7 \div 2$ or 1.35 milliamperes per volt, (milli-siemens).

The amplification factor. Derivation of μ

21. Since $\mu = g_m \times r_a$, the amplification factor of this valve can now be determined. It is

$$g_m \times r_a = \frac{1\cdot35 \text{ amperes.}}{1000 \text{ volts}} \times 10,000 \frac{\text{volts}}{\text{amperes}} = 13\cdot5.$$

The significance of the amplification factor must be fully realised, since it is its amplifying property which gives the triode such a predominant importance in modern radio technique. It may be expressed by stating that in a valve of amplification factor μ one unit change of voltage between grid and filament will cause the same change of anode current as a change of μ units of voltage in the P.D. between anode and cathode. Referring again to fig. 12, with an anode potential of 100 volts and a grid potential of +1 volt, the anode current is 7.2 milliamperes. A reduction of anode potential by 20 volts would reduce this current to 5 milliamperes. If, however, the anode potential were kept constant at 100 volts, the same reduction of anode current would be achieved by the application of a potential of $-0\cdot5$ volts to the grid, a change of 1.5 volts. From these figures the grid potential is approximately 13 times as effective as the anode potential in producing a change of anode current. The slight variation in the value of μ , as calculated by different methods, is of no practical significance.

The extent to which the anode current is changed by a given change of grid voltage will depend upon the extent to which the grid screens the plate from the electron flow, and also upon the degree to which the grid voltage influences the potential gradient in the space between cathode and anode. Both these factors depend upon the shape and disposition of the electrodes and therefore the amplification factor μ depends upon the geometry of the valve. This factor is large if the grid is situated comparatively near to the cathode and has a fine mesh, while a coarse

CHAPTER VIII.—PARA. 22

grid near to the plate results in a low value for μ . In fig. 13 the alternative locations and wire diameters shown at A and B respectively will give approximately the same amplification factor.

The qualification "approximately" has been inserted because in the figure referred to it is implicitly assumed that electrons travel from filament to anode in straight lines. Fig. 14 has been developed to give a rather more accurate representation of the actual flow of electrons. The thin solid lines in this figure show the direction of the electrostatic field between anode and filament in the absence of any emission from the latter, for three different values of grid potential.

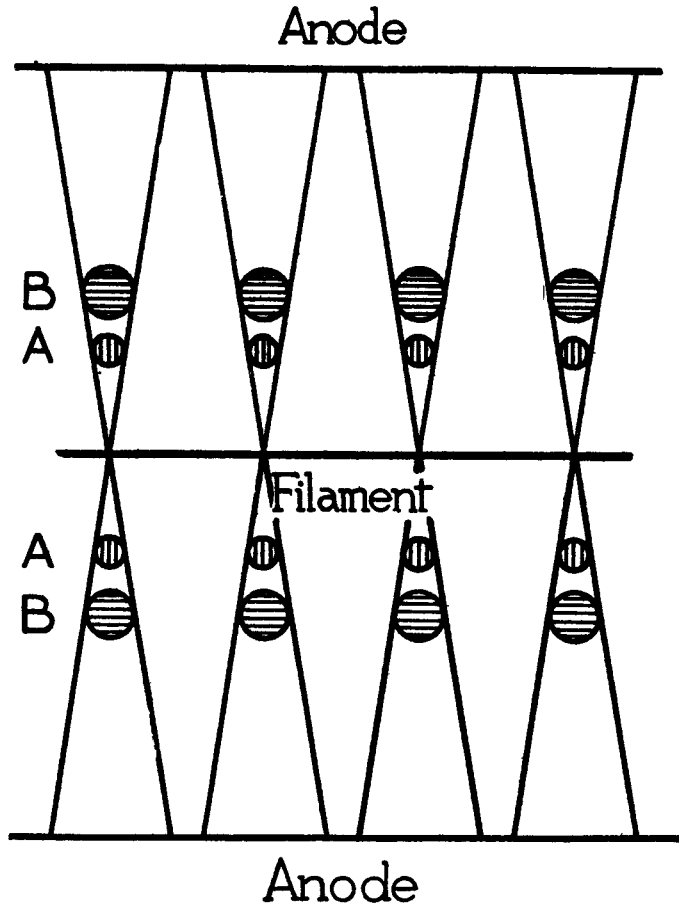
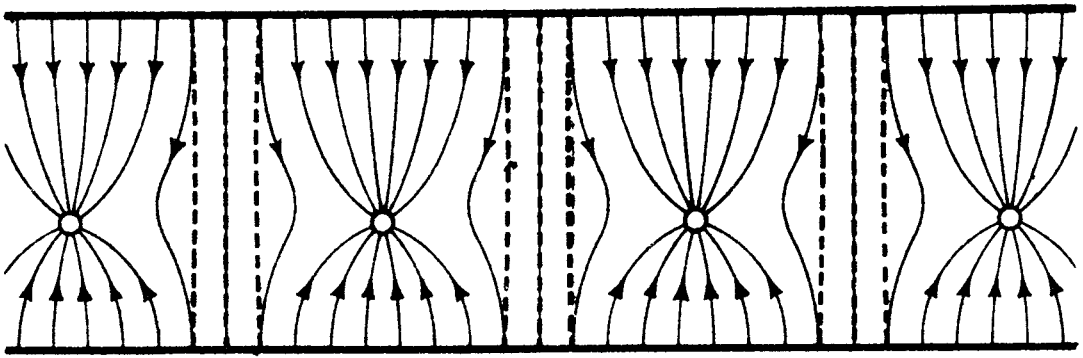


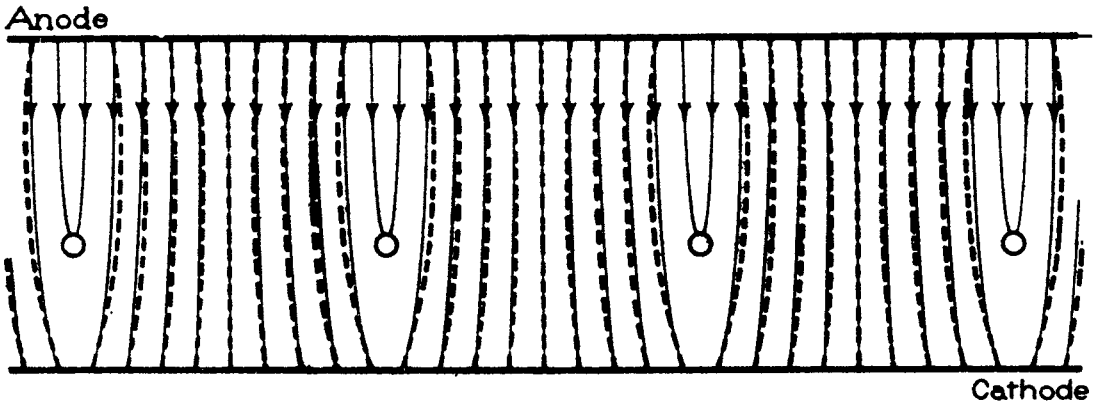
FIG. 13, CHAP. VIII.—Alternative designs of grid giving approximately equal amplification factor.

The heavy broken lines represent the paths of electrons when emission is taking place, and as the electrons flow along lines of electric force, the broken lines may also be assumed to represent the electric field under emission conditions. The effect of a positive grid potential in increasing the electron current is also shown by this figure.

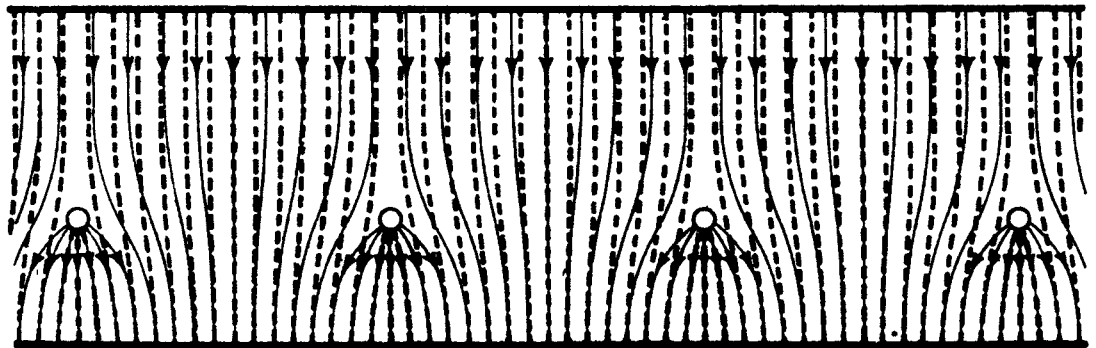
22. The principal function of the triode is therefore to act as an extremely sensitive "relay." It differs from an electro-mechanical relay in two particulars. First, it has no appreciable time lag, the change of anode current taking place almost instantaneously upon the occurrence of a grid voltage change. This is due to the almost entire absence of mass, and therefore inertia,



A - Grid Negative



B - Grid Neutral



C - Grid Positive

ELECTRIC FIELDS IN TRIODE

in the moving part of the relay, the electrons themselves. Second, by suitable adjustments of the operating conditions, the valve can be made to perform its amplifying function without constituting a load upon the preceding circuit, the presence or absence of the valve and its anode circuit making no difference to the power consumption in the circuits normally connected to grid and filament. It might at first sight be thought that this is contrary to the principle of the conservation of energy, because energy is dissipated in the valve itself, but the anode circuit contains a source of energy (the H.T. supply device), and it must be realised that it is this source which supplies the energy dissipated in the valve and the anode circuit.

The $I_g - V_g$ characteristic

23. The principal function of the grid is to act as a controller of the flow of anode current, but under certain conditions it may collect electrons, while if an appreciable quantity of gas is present in the valve, it may sometimes collect positive ions. A typical $I_g - V_g$ characteristic curve for a soft triode is shown in fig. 15. It must be clearly understood that this is not the valve

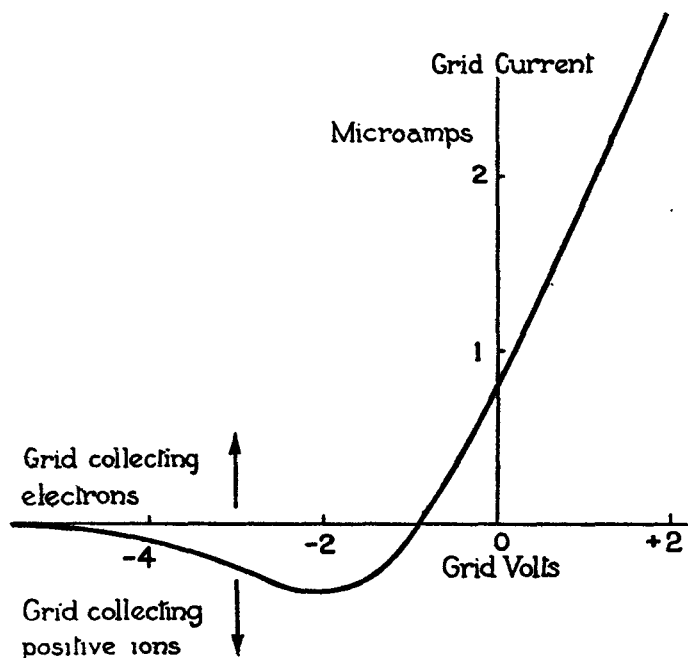


FIG. 15, CHAP. VIII.—Grid current/grid volts characteristic of soft triode.

whose $I_a - V_g$ and $I_a - V_a$ curves have already been discussed. The general features of this curve may be explained as follows. When electrons flow through a valve, they travel with accelerated velocity, under the influence of the electric field of force set up by the H.T. battery. If gas molecules are present, collisions occur between them and the electrons, resulting in ionisation of the gas. Whether any one collision will result in the formation of a positive or a negative ion depends largely upon the velocity of the electron, or how far it has travelled since leaving the cathode before meeting the molecule, the average distance for the whole emission being called the mean free path of any electron. If the velocity of the electron on impact is low, the electron probably unites with the molecule forming a negative ion, while if it is high, the impact may cause an electron to be dislodged from the molecule, resulting in the formation of a positive ion. Negative ions move off toward the anode with comparatively low velocity, while positive ions move toward the filament. The formation of negative ions thus results in a reduction of anode

CHAPTER VIII.—PARAS. 24-25

current, and the formation of positive ions in an increase. So far we have only considered the effect of the gas on the anode current. Now consider the grid. If its potential with respect to cathode is positive, the grid will collect electrons from the anode stream, while if negative it will collect positive ions provided that these are formed. If the grid is extremely negative, its repulsive force will prevent any electrons travelling an appreciable distance from the cathode, and therefore the production of positive ions is unlikely. The collection of positive ions by the grid is referred to as reverse grid current.

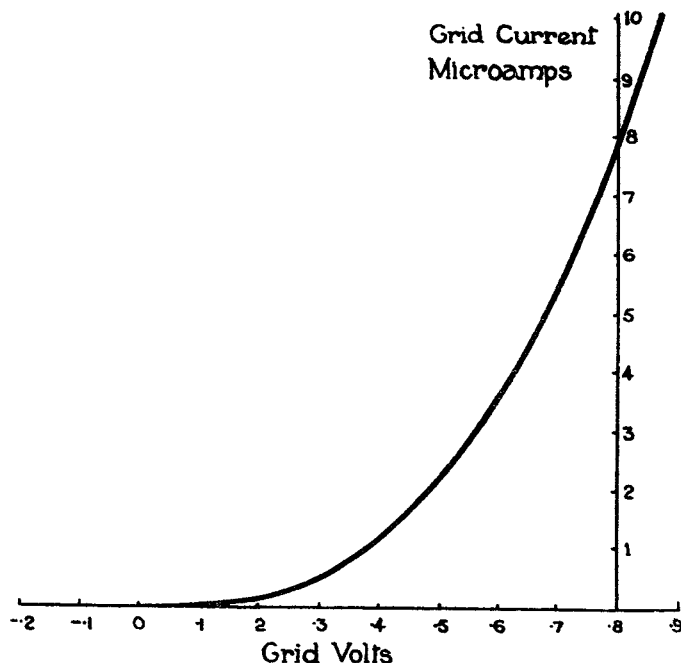


FIG. 16, CHAP. VIII.—Grid current/grid volts characteristic of Valve V.R. 21.

The $I_g - V_g$ curve for a perfectly hard triode would show no reverse grid current because no positive ions would be formed. Since a perfect vacuum is unattainable, it is usual to specify the degree of vacuum in terms of the permissible reverse grid current, a typical specification for a receiving valve being "maximum reverse grid current not to exceed two microamperes."

The $I_g - V_g$ curve of the valve previously discussed is shown in fig. 16. It is a hard valve, the reverse grid current being indistinguishable on the scale to which this curve is plotted.

The $I_g - V_a$ characteristic

24. This curve is rarely used, and it has only been mentioned in order to emphasise that the triode has four characteristics. A valve having n electrodes has $(n - 1)^2$ characteristics, but its behaviour in most circumstances can be deduced from two of them.

Simultaneous variation of V_a and V_g — dynamic characteristics

25. So far we have considered the variation of anode current under conditions in which either V_a or V_g are varied independently. In practical circuits however, V_a and V_g usually vary simultaneously. In the valve characteristic circuit, fig. 10, the resistance of the external anode circuit is utterly negligible compared with the internal resistance of the valve, but in actual

working conditions the anode circuit invariably contains an impedance of some kind, and this will modify the performance of the valve. We will first assume that the anode circuit possesses a resistance of the same order as the anode A.C. resistance of the valve.

The triode with its anode load resistance, as it is termed, is shown diagrammatically in fig. 17 in which for purposes of numerical illustration the valve may be the V.R. 21, the static characteristics of which are given in figs. 11, 12 and 16. The external anode circuit consists of a 120 volt battery and a non-inductive resistance R of 10,000 ohms. The grid may first be assumed to have zero potential with respect to the filament. On completing all circuits an anode current will be established, setting up a P.D. across the resistance. The P.D. between anode and filament will then be, not 120 volts, but 120 volts minus the fall of P.D. in the resistance. The effect of the latter is therefore to reduce the anode current, an effect which might be anticipated from first principles, but nevertheless is often not appreciated.

Now suppose the grid is given a positive potential. There will be an increase in the anode current, and consequently a larger P.D. between the terminals of the load resistance. The anode-filament P.D. being equal to the battery E.M.F. minus this $I R$ drop, the rise of anode current will be accompanied by a fall in the anode-filament P.D. On the other hand, if the grid is given a negative potential, the electron flow decreases, and the P.D. between the terminals of the load resistance will be less than when the grid potential was zero. The anode-filament P.D. therefore rises as the grid is made more negative with respect to the filament. It follows therefore

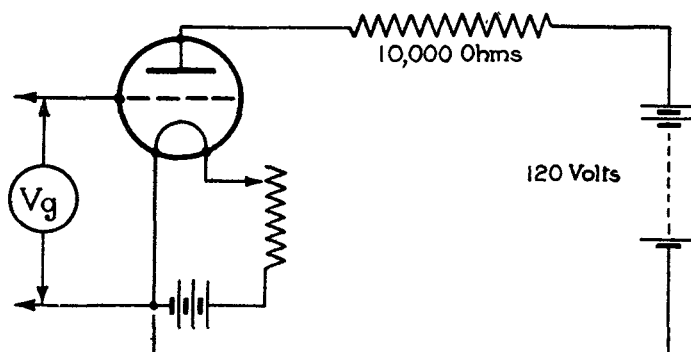


FIG. 17, CHAP. VIII.—Circuit illustrating derivation of dynamic characteristic.

that if an alternating voltage is applied between grid and filament, the changes of anode current and anode-filament P.D. are in anti-phase, an increase of I_a due to the positive half-cycle of grid voltage being accompanied by a decrease in V_a and vice-versa. Algebraically,

$$\delta V_a = - R \delta I_a$$

the minus sign being inserted to denote the anti-phase relation. The curve showing the variation of anode current with variation of grid voltage, for a given load impedance—not necessarily non-reactive—is called the dynamic mutual characteristic.

26. For a given load resistance, the dynamic mutual characteristic can be derived from the family of static mutual characteristics as follows. Referring to figs. 12 and 17, the anode current is just reduced to zero by the application of -9.5 volts to the grid. The P.D. across the load resistance is also zero, and the anode-filament P.D. equal to the E.M.F. of the H.T. battery, 120 volts. The point $I_a = 0$ on the 120 volt curve is therefore a point on the dynamic characteristic. If the negative grid voltage is reduced so that the anode current rises to 2 milliamperes, the P.D. across the load is 20 volts, and the anode-filament P.D. falls to $120 - 20$ or 100 volts. The point corresponding to 2 milliamperes on the 100 volt curve therefore gives another point on the dynamic characteristic. A further reduction of negative grid voltage, allowing the anode current to rise to 4 milliamperes, causes a P.D. of 40 volts across the load resistance and the

CHAPTER VIII.—PARA. 26

anode-filament P.D. falls to $120 - 40 = 80$ volts; the point corresponding to 4 milliamperes on the 80 volts curve gives a third point on the dynamic characteristic. These points are indicated by small circles in fig. 12 and the dynamic characteristic for a resistance of 10,000 ohms may now be drawn. To avoid confusion it has been shown in a separate diagram, fig. 18.

The slope of this curve is the dynamic mutual conductance for the given load resistance, and is measured in milliamperes per volt as in the case of the mutual characteristic. The slope may be found algebraically as follows. Since $g_m = \frac{\delta I_a}{\delta V_g}$ the change of anode current δI_a for any change of grid voltage δV_g , is $g_m \delta V_g$ (provided V_a is constant) and in the same manner $\delta I_a = g_a \delta V_a$ provided V_g is constant. If both V_a and V_g vary simultaneously, the total variation of I_a is given by the sum of the separate variations, so that

$$\delta I_a = g_m \delta V_g + g_a \delta V_a$$

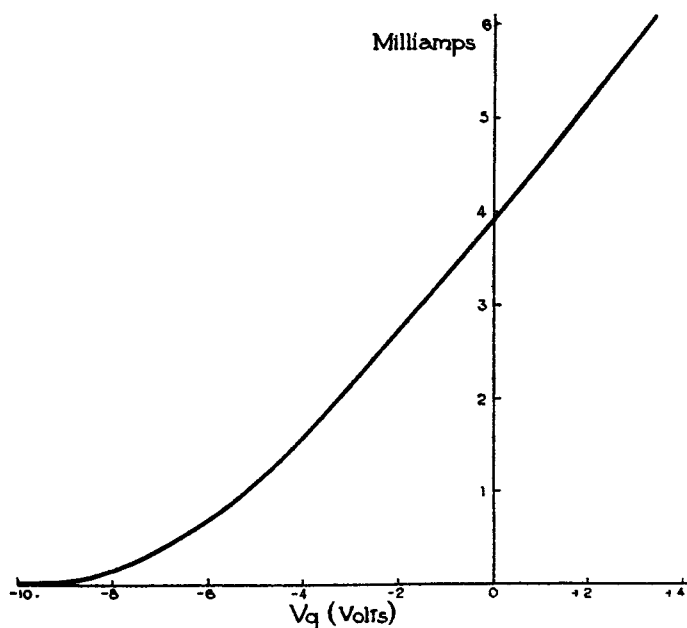


FIG. 18, CHAP. VIII.—Dynamic characteristic of Valve V.R. 21 with 10,000 ohms resistance.

bearing in mind that g_m and g_a are only constant within the region of linear operation. It is often more convenient to use a different notation, in which i_a is written for δI_a , v_a for δV_a and v_g for δV_g . In this form

$$i_a = g_m v_g + g_a v_a$$

In the foregoing discussion, it was shown that, under dynamic conditions, $\delta V_a = -R \delta I_a$ which in the new notation may be combined with the last equation, giving

$$i_a = g_m v_g - g_a R i_a$$

$$(1 + g_a R) i_a = g_m v_g$$

$$\therefore i_a = \frac{g_m}{1 + g_a R} v_g$$

Thus $\frac{g_m}{1 + g_a R}$ or $\frac{g_m}{1 + \frac{R}{r_a}}$ is the slope of the dynamic mutual characteristic, just as g_m is the slope of the static mutual characteristic.

It is often convenient to write this relation in a different way. Since

$$i_a = \frac{g_m v_g}{1 + \frac{R}{r_a}} = \frac{r_a g_m v_g}{r_a + R}$$

$$i_a = \frac{\mu v_g}{r_a + R}$$

because $r_a g_m = \mu$

27. It has already been stated that the fundamental function of the triode is that of an amplifier, and a somewhat detailed consideration of the factors governing its employment in this capacity is given in Chapter XI. The triode is also employed as a power converter and as a rectifier in connection with radio transmitters and receivers respectively; and it is necessary to give a brief outline of the use of the valve as an amplifier before the latter applications can be appreciated. Suppose the steady anode voltage and mean grid voltage are so adjusted that the

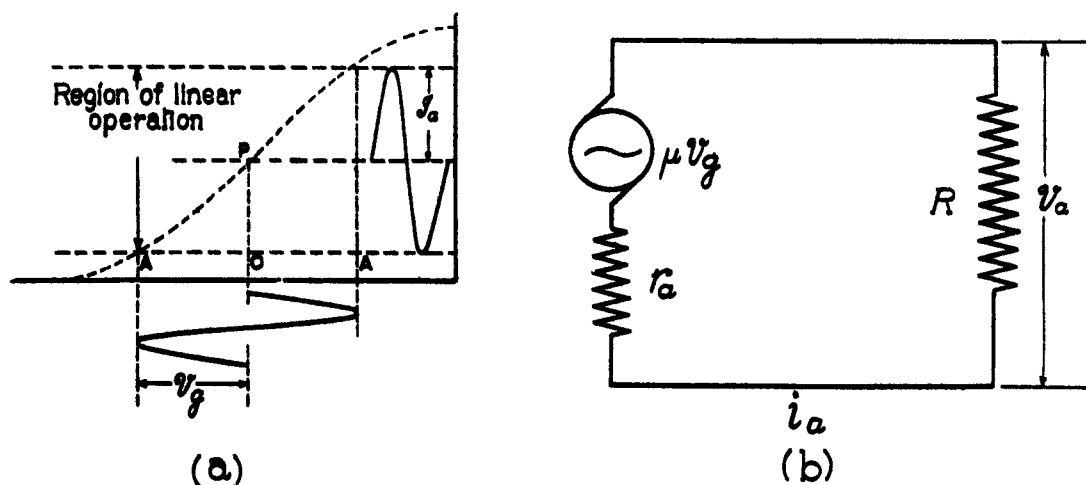


FIG. 19, CHAP. VIII.—Triode used as amplifier; operating conditions and equivalent circuit.

anode current is equal to one-half the saturation value, the mean anode current being that marked P on the dynamic characteristic fig. 19a. If now an alternating voltage $v_g = \mathcal{V}_g \sin \omega t$ is applied between grid and filament, the peak value of which does not exceed O A, a sinusoidal variation of anode current will take place as shown in the diagram.

By definition of the amplification factor the voltage will produce a change of anode current equal to that which would be produced by μ times this voltage acting directly in the anode circuit, provided that the variation of grid voltage is confined within the limits of linear operation. Within this limit therefore, the valve may be considered, for purposes of calculation, to act as the generator of an alternating E.M.F. μv_g , and to possess an internal resistance r_a ohms. The valve with its associated anode circuit may be represented by the equivalent circuit fig. 19b, the anode circuit load being assumed to consist only of a resistance R .

The variation in anode current due to the sinusoidal grid-filament voltage will be

$$i = \frac{\mu \mathcal{V}_g \sin \omega t}{r_a + R}$$

CHAPTER VIII.—PARA. 28

which is greater than $\frac{V_g \sin \omega t}{R}$ provided that μ is greater than $1 + \frac{r_a}{R}$. The latter condition may be fulfilled by the choice of a suitable valve for R , or, if this is fixed, by choice of a valve having suitable values of μ and r_a . The valve may thus be considered to act as a current amplifier.

The P.D. set up between the ends of the external resistances as a result of the change of grid voltage v_a is Ri_a and

$$Ri_a = \frac{R \mu v_g}{r_a + R}$$

Now — Ri_a is the change of anode-filament P.D. and may be denoted by v_a as before, hence

$$v_a = - \frac{R}{r_a + R} \mu V_g \sin \omega t$$

Thus v_a is a sinusoidal voltage variation also. It is again seen that v_a is larger than v_g if $\mu > 1 + \frac{r_a}{R}$ and the valve is said to function as a voltage amplifier. When the mean anode voltage V_a and mean grid voltage V_g are so chosen that the mean operating point lies near the upper or lower bends of the $I_a - V_g$ curve, the variation of anode current will not be a true reproduction of the grid voltage variation. These are the conditions in which a valve is employed as a rectifier in radio reception, and in these circumstances it is no longer permissible to employ the simple equivalent circuit for purposes of calculation.

The load line

28. Dynamic conditions can frequently be studied with greater facility with the aid of the $I_a - V_a$ characteristics. The procedure is similar to that adopted in deriving the dynamic mutual characteristic. Referring to figs. 12 and 17 we have seen that with 120 volts H.T. supply, a negative grid bias of about 9.5 volts will reduce the anode current to zero, and the anode-filament P.D. will then be 120 volts. Also, on decreasing the negative bias until I_a rises to 2 milliamperes the P.D. across the load resistance is 20 volts and the anode-filament P.D. 100 volts. A further reduction of the grid bias to zero causes the anode current to rise to 4 milliamperes, and the anode-filament P.D. falls to 80 volts. These points are plotted in fig. 11, and the straight line passing through the points is called the load line corresponding to an anode load resistance of 10,000 ohms.

When an alternating P.D. is superimposed upon the steady grid bias voltage a point representing corresponding instantaneous values of anode current and anode-filament P.D. travels to and fro on the appropriate load line, and the chief advantage of this method of representation is the ease with which the power relations can be computed. First, take the conditions in the absence of the alternating grid-filament voltage. The $I_a - V_a$ curves in fig. 20 are typical of those of a small power triode; let the H.T. supply voltage be 120 volts and the load resistance be 10,000 ohms as before; the corresponding load line has been inserted in the diagram. If the grid bias is fixed at —8 volts, the anode current I_a is 5 milliamperes and the power supplied by the H.T. battery, 600 milli-watts. The anode-filament P.D. V_a is only 70 volts, and the power dissipated in the valve is $I_a V_a = 70 \text{ volts} \times 5 \text{ milliamperes}$ or 350 milli-watts. The power dissipated by the load resistance is $I_a^2 R$ or 250 milli-watts.

Now suppose an alternating voltage of 8 volts peak value is applied between grid and filament in addition to the steady bias. During the positive half-cycles of grid voltage the anode current will increase, rising to a maximum value of 8 milliamperes, and will fall during negative half-cycles, its minimum value being only 2 milliamperes. A point representing corresponding instantaneous values of anode current and anode filament P.D. therefore travels to and fro between the points A and B. (fig. 20). Owing to the presence of this alternating component of anode current in the resistance R , the power losses in the latter will be increased. If $\mathcal{I}_a$ is the peak value of the alternating component, its R.M.S. value is $\frac{\mathcal{I}_a}{\sqrt{2}} = \frac{3.0}{\sqrt{2}}$ milliamperes and the

power dissipated in the load resistance is $\frac{I_a^2 R}{2} = 45$ milli-watts. This of course is in addition to the amount calculated above, due to the steady anode current.

The average value of the current, however, is still 5 milliamperes, and the power delivered by the H.T. battery is unchanged. The power expended in the valve is the difference between the total power supplied and that dissipated in the load resistance, and this is less than before, i.e. 350-45 or 305 milli-watts instead of 350 milli-watts.

In fig. 20, O F is the mean anode current I_a ; O C the H.T. battery voltage E_a ; O G the anode-filament P.D. V_a ; A Q the peak value of the alternating component of anode current, I_a ; P Q the peak value of the alternating voltage across the anode load resistance. Then O G = $I_a r_a$, G C = $I_a R$ and the rectangle O C D F = $I_a^2 (R + r_a)$ and is equal in area to

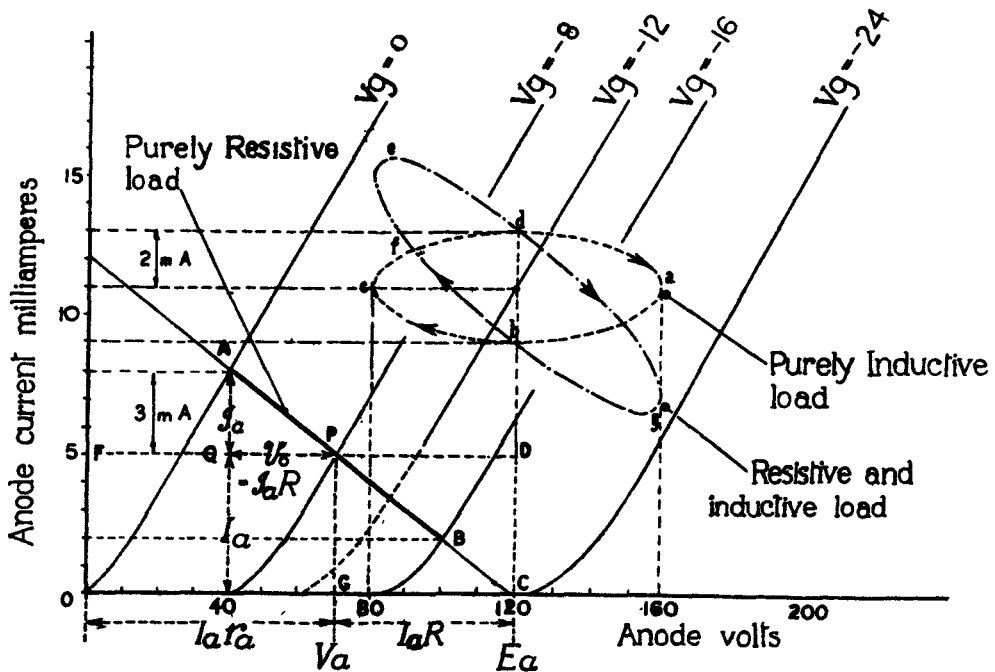


FIG 20, CHAP. VIII.— $I_a - V_a$ characteristics of small power triode, showing load lines for resistance, reactance and impedance loads.

the power supplied by the H.T. battery. The area G C D P is equal to the steady power dissipation $I_a^2 R$ in the resistance R. The area O G P F is equal to the power dissipated in the valve itself, $I_a^2 r_a$, and the triangular area P Q A to the alternating power dissipated in the load resistance = $\frac{I_a^2 R}{2}$.

29. If the anode circuit contains a purely reactive impedance, e.g. an inductive coil of negligible resistance, the change of P.D. corresponding to a given change of anode current will lag on the latter by 90° and the load line becomes an ellipse upon which the representative point travels in a clockwise direction once per cycle. Taking the same valve as before and referring to fig. 20 let the H.T. supply voltage be 120 volts and the mean grid bias—12 volts. If the anode load is an inductive reactance of 20,000 ohms and the alternating grid voltage has a peak value sufficient to cause a P.D. of 40 volts between the load terminals, the peak value of the alternating

component of anode current will be 2 milliamperes. Since the P.D. lags by 90° on the current change, the anode current will be of normal value, 11 milliamperes, at the instant when the anode-filament P.D. is 40 volts above normal, i.e. 160 volts, giving the point *a* on the load ellipse. At the instant when the anode-filament P.D. is normal the anode current will be 2 milliamperes below normal, i.e. 9 milliamperes, giving the second point *b*. Similarly, when the P.D. falls to 80 volts the anode current will rise to normal, giving the point *c*, and finally when the anode-filament P.D. reaches normal again, the anode current will reach its maximum value during the cycle, namely 13 milliamperes, point *d*. The complete elliptical load line thus traced but is shown by a dotted line in the diagram.

When the anode impedance possesses both resistive and reactive components, the load line is an ellipse inclined to the vertical, its mean slope corresponding to the effective resistance. Suppose the operating conditions to be as in the last instance, except that a resistance of 10,000 ohms is connected in parallel with the inductive load. The load line is then found by taking the

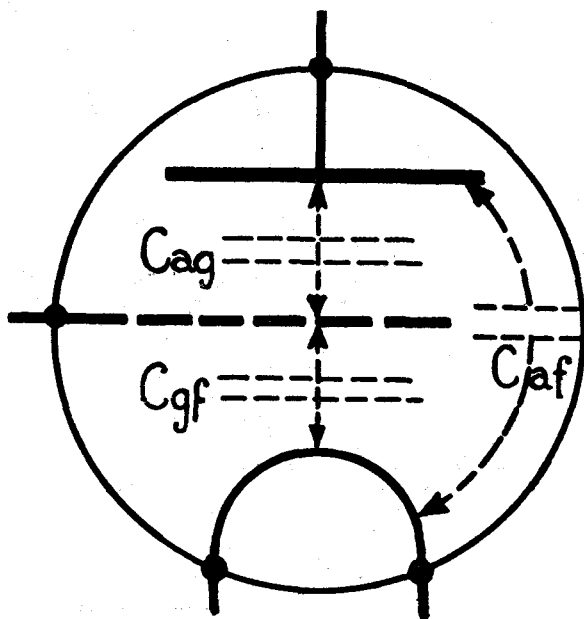
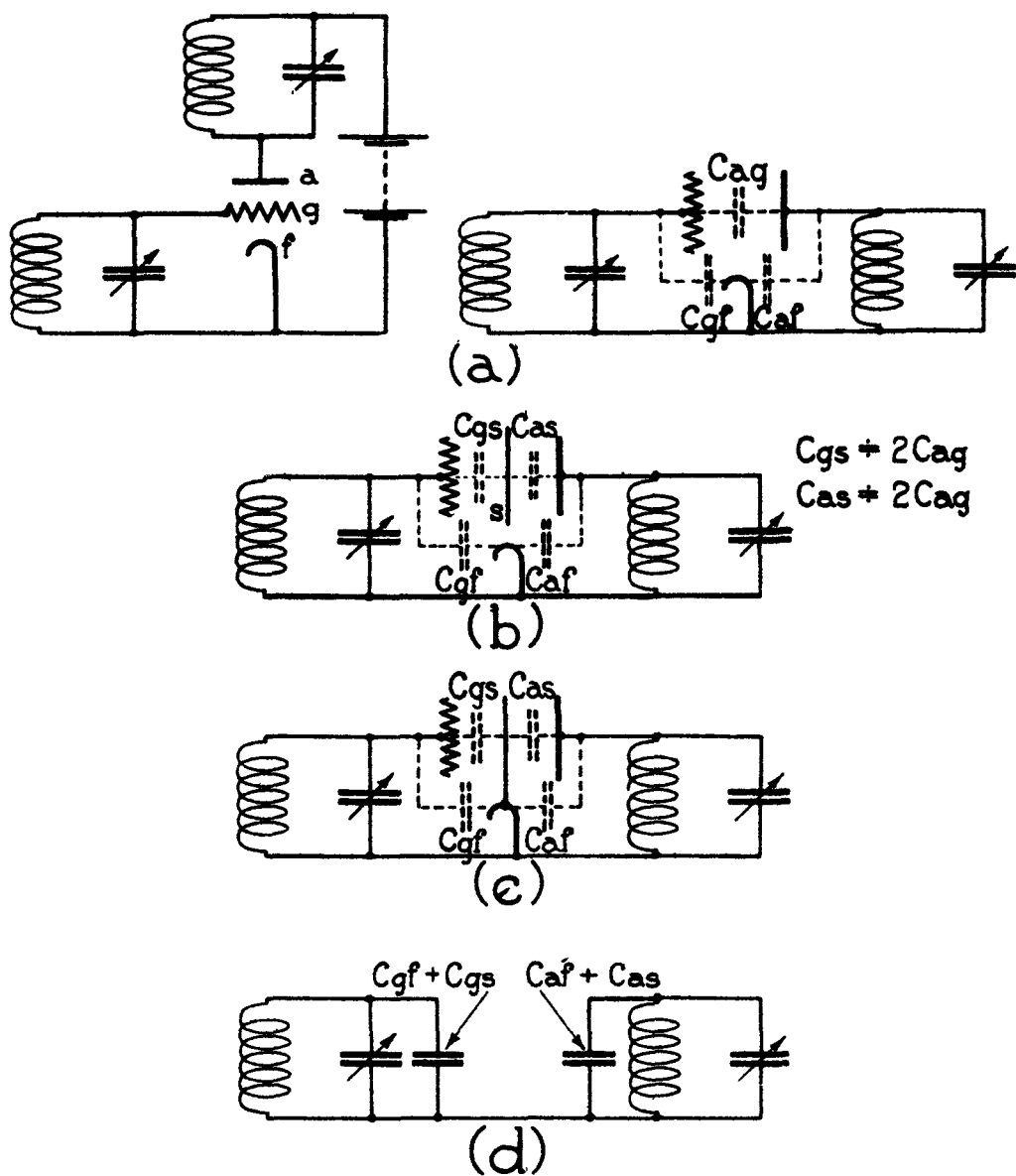


FIG. 21, CHAP. VIII.—Diagrammatic representation of inter-electrode capacitance.

algebraic sum of the currents in the two branches of the external impedance during the whole cycle, giving the ellipse *e, d, g, b, f* (fig. 20). It is however seldom necessary to consider load lines of this nature.

The dissipation line

30. The application of an anode-filament potential and consequent flow of anode current results in a loss of power due to the heat developed in the valve, and as previously stated transmitting and power amplifying valves are rated according to the power they are capable of dissipating in this manner. The power dissipated (in milli-watts) is given by the expression $P = V_a I_a$, where V_a and I_a are in volts and milliamperes respectively. For any given valve, the permissible dissipation is constant and the equation represents a rectangular hyperbola, which may be drawn on the $I_a - V_a$ characteristics as follows. Taking the valve, receiving, V.R.21 as an example, and assuming that it is capable of dissipating 200 milli-watts, the following co-ordinates viz. (10 m.a., 20 volts), (5 m.a., 40 volts), (2.5 m.a., 80 volts), (2 m.a., 100 volts), (1.25 m.a., 160 volts),



EFFECT OF SCREEN BETWEEN GRID AND ANODE

(1 m.a., 200 volts), etc., are all points on the 200 milli-watt dissipation line, which has been plotted from these values in fig. 11. The mean anode current under working conditions must lie on or below this line, otherwise the power dissipated by the valve will be in excess of the rated power. It is not customary to give the power rating of a receiving valve, and the foregoing assumption of 200 milli-watts for the valve V.R. 21 was made purely to demonstrate the method of drawing the curve.

Inter-electrode capacitance of the triode

31. The electrodes of a valve and the connecting leads to the external circuit are in somewhat close proximity, and therefore each pair of electrodes possesses a capacitance which although comparatively small may have a profound influence upon the behaviour of the valve, particularly when used at the higher radio-frequencies. Fig. 21 indicates the notation generally adopted in writing of these capacitances, the usual magnitudes of each being of the order of from 5 to 10 micromicrofarads in ordinary triodes. The inter-electrode capacitance is never absolutely constant since the space between the electrodes contains free electrons, which influence the effective dielectric constant. The most important inter-electrode capacitance is that denoted by C_{ag} in fig. 21 because in practice, the alternating grid-filament P.D. is often that developed between the terminals of the inductance of a tuned circuit, while the anode load impedance consists of a parallel-resonant circuit tuned to the same frequency. The anode-grid capacitance C_{ag} then takes the place of the coupling condenser in a well-known form of capacitance-coupled circuit (Chapter VI), the actual circuit and its electrical equivalent being shown in fig. 22a. The complications arising from this coupling in the case of radio-frequency amplifiers are dealt with in Chapter XI, but it may here be stated that the effect is to limit severely the amplification obtainable at frequencies above about 1,000 k.c/s. Many attempts have been made to overcome this difficulty by special circuit arrangements, but these cannot be considered entirely satisfactory for use in receivers, in which the circuit adjustments must often be changed with rapidity and accuracy. The screen-grid valve and its later development the radio-frequency pentode were evolved in a successful attempt to attack the problem at its source, by a reduction in the effective value of the coupling capacitance.

SCREEN-GRID AND PENTODE VALVES

The screen-grid valve

32. This type of valve has four electrodes and is therefore sometimes referred to as a "tetrode." The electrodes consist of a cathode or electron emitter, control grid, screening grid and anode or collecting electrode. The function of the control grid is exactly the same as in the triode, namely, to control the flow of electrons so that the valve will act as a relay. The screening grid acts as an electrostatic screen between the control grid and the anode, thus effecting a considerable reduction in the effective grid-anode capacitance referred to in the preceding paragraph. A consideration of fig. 22 will make this clear; at (a) the tuned input and output circuits of a triode are shown, coupled together by the inter-electrode capacitance C_{ag} . If a metal plate (s) is interposed between grid and anode as at (b), one condenser C_{gs} is formed by the grid (g) and plate (s) and another, C_{as} , by the plate (s) and anode (a), the capacitance of each being larger than the original capacitance C_{ag} , but as the two are in series between grid and anode, the total effective capacitance of these electrodes is unaltered. The coupling effect of the grid-anode capacitance is therefore not affected by an insulated screen. If the screen is connected to the filament (f) as in (c), the condenser C_{gs} formed by the grid and screen is in parallel with the capacitance C_{gf} while the condenser C_{as} formed by the screen and the anode is in parallel with the capacitance C_{af} (fig. 22d). There is therefore no effective capacitance whatever between the grid and the anode and consequently the input and output circuits are not coupled together, that is, energy can no longer be transferred from one to the other. It must be borne in mind that a single connecting conductor does not constitute coupling in the electrical sense.

In practice the screening cannot be perfect, since it is necessary to use a screen in the form of a gauze or mesh in order that electrons may pass through it on their way from filament to anode. Some lines of electric force from the anode inevitably terminate upon the grid, and

CHAPTER VIII.—PARA. 33

hence there must be some residual anode-grid capacitance. By correct design this can be reduced to a value about one-thousandth of the anode-grid capacitance of a triode having similar dimensions.

33. The introduction of such a screen directly connected to the filament would modify the characteristics of the valve in such a manner as to render it unsuitable for practical use, in particular the anode A.C. resistance would be extremely high. This disability is overcome by the application of a positive potential to the screening electrode, its value being generally variable within the limits of one half to five-eighths that of the anode potential. This does not nullify the

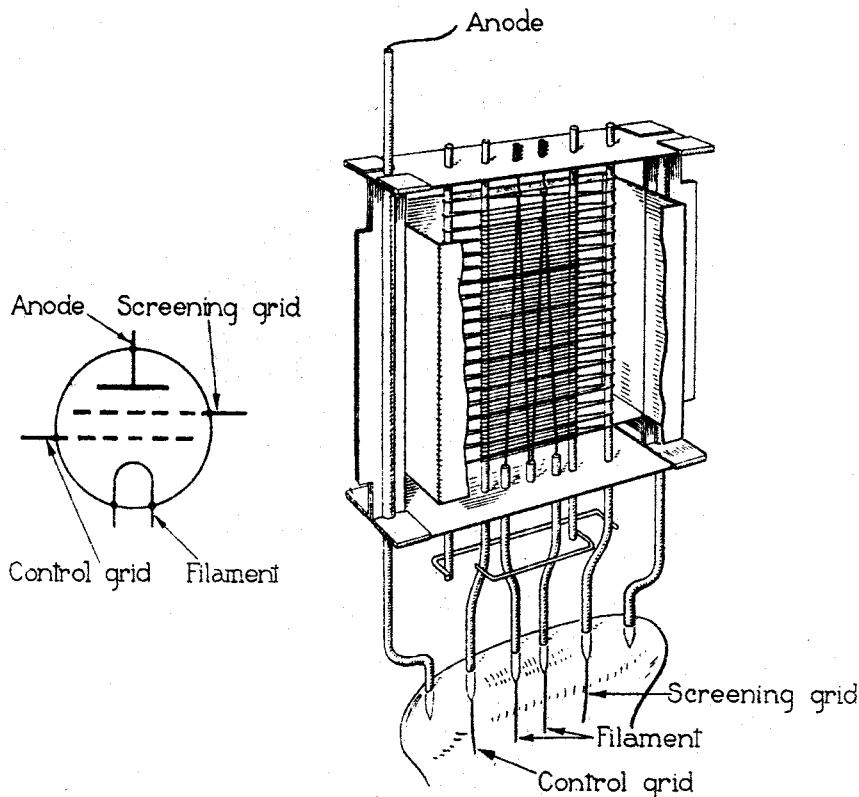


FIG. 23, CHAP. VIII.—Electrodes of screen-grid valve.

screening properties, provided that the screen is connected to the cathode by a low impedance path, and it is usual to ensure this low impedance (for the frequency at which the valve is to function) by connecting externally a condenser of about $\cdot 5$ microfarad between screen and filament terminals.

The appearance of the electrodes of a typical screen-grid valve with directly heated cathode is shown in fig. 23 in which also the external connections are indicated. It should be noted that the anode is connected to the top terminal, the pin connection which serves as anode connection in a triode being allotted to the screen. In order that the screening may be complete, the bulb is generally metallised, the metallising being connected to the negative filament pin as usual. It is also necessary to ensure that the input and output circuits are effectively screened from each other, otherwise the object of the valve is defeated.

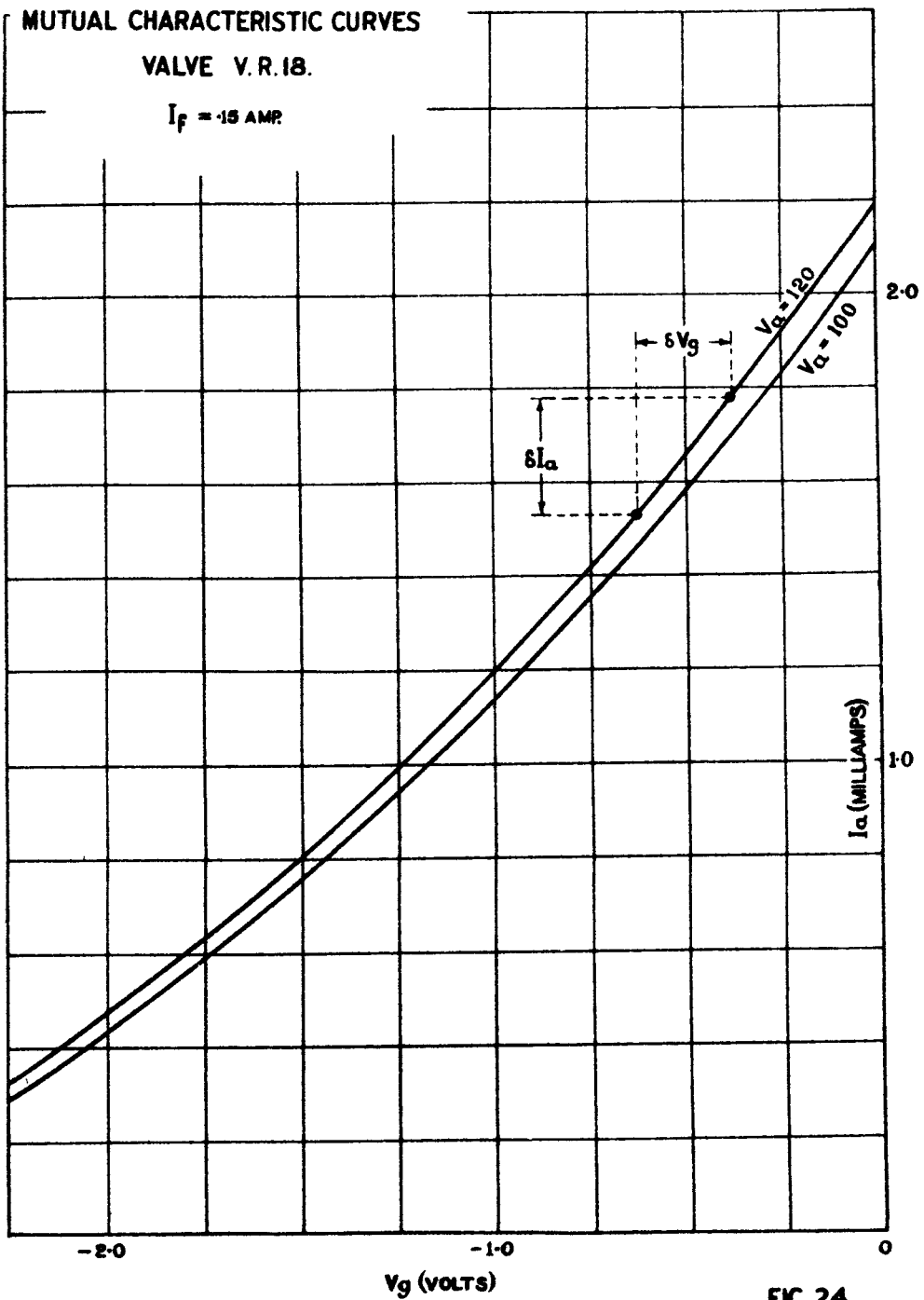
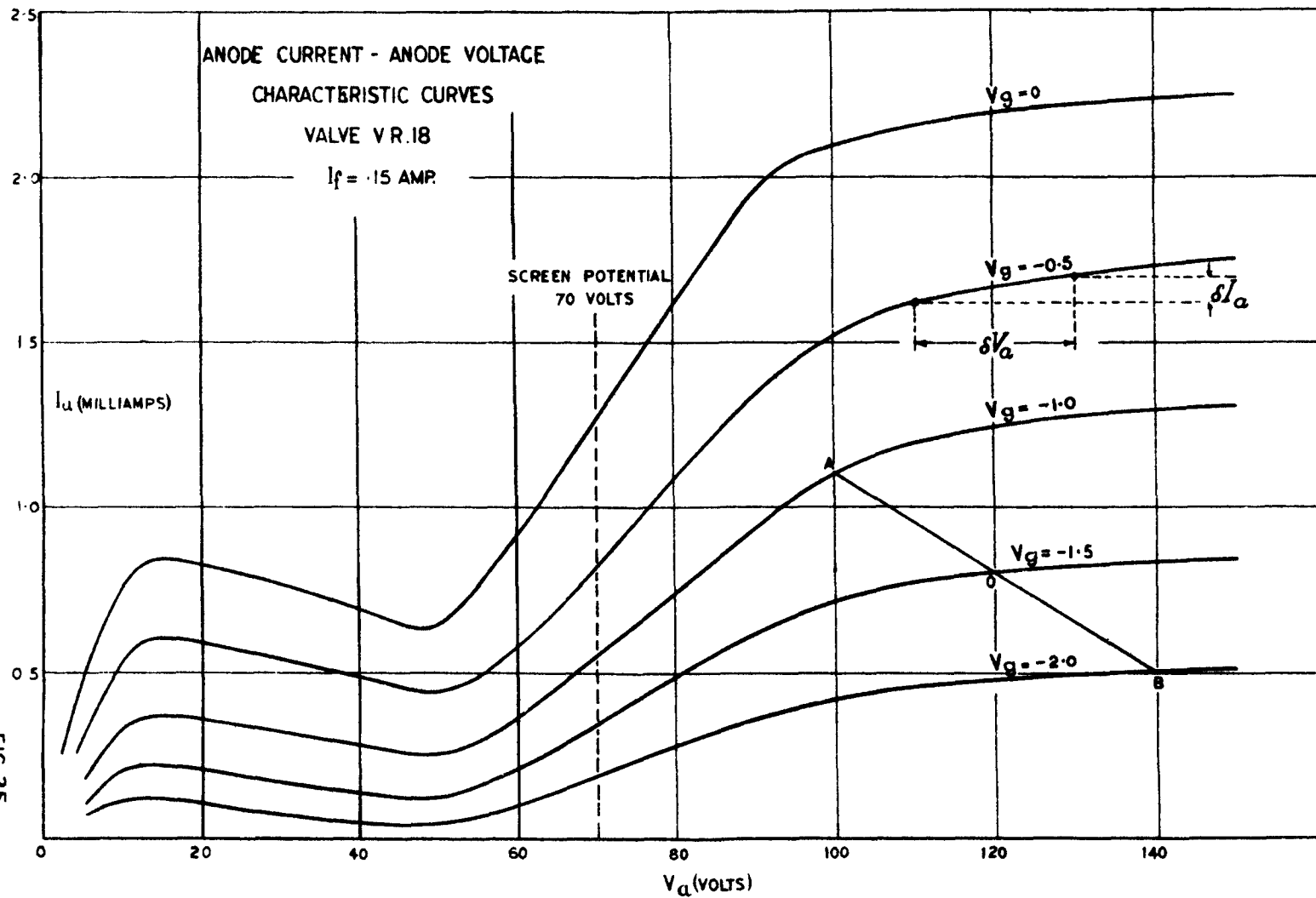
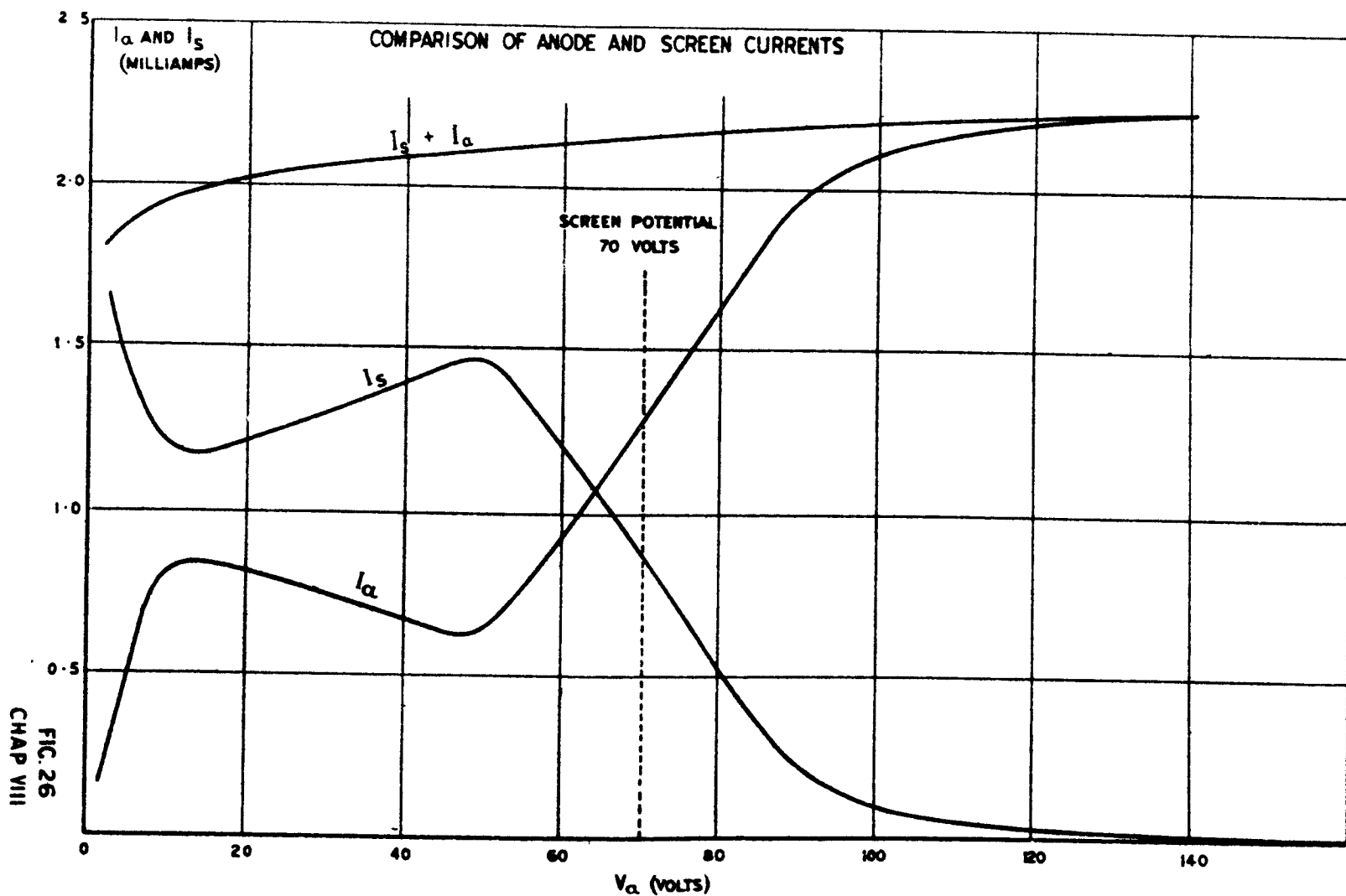


FIG. 24
CHAP VIII

FIG 25
CHAP VIII





Characteristic curves

34. The static mutual characteristics of a service S.G. valve (V.R.18) are given in fig. 24. It will be seen that they are similar to those of a triode in general character. As the mutual conductance of any valve of given emission depends chiefly on the design of the control grid, tetrodes and triodes with identical cathodes and control grids have similar if not identical values of mutual conductance. The individual curves of the $I_a - V_g$ family lie much closer together for the tetrode than for the triode which implies a higher anode A.C. resistance. This feature is due to the influence of the positive potential of the screen in assisting the production of an electron current. Slight variations of the anode potential make very little difference to the anode current because practically no lines of electric flux extend from anode to cathode, owing to the effective screening between the two electrodes. The point may be illustrated by reference to the anode current-anode voltage curves fig. 25 in which it can be seen that provided the anode potential is at least 30 volts above that of the screen an increase of anode voltage produces only a very small increase of anode current. The value of r_a for this particular valve calculated from the $I_a - V_a$ curve is 266,000 while the value of g_m (from the mutual characteristic) is 1 milliampere per volt. The amplification factor is 266, which is much greater than that of a triode.

35. Another notable feature of the $I_a - V_a$ curves is the region of negative slope when the anode potential is below that of the screen. Fig. 26 shows the anode current, screen current and their sum, plotted against anode potential, the screen potential being fixed at 70 volts. When the anode potential is zero, all the electrons passing the control grid go to the screen, and the screen current is correspondingly large. On raising the anode potential some of the electrons pass through the screen to the anode. The total number of electrons flowing is not greatly increased, and the rise of anode current is largely at the expense of the screen current, with a resultant decrease of the latter. When the anode potential exceeds 15 volts the anode current starts to decrease with an increase of anode voltage, the screen current increasing proportionally. This phenomenon is due to the emission of electrons from the anode.

In the opening paragraphs of this chapter secondary emission was described as the production of emission from a body by bombardment with electrons. In the region under consideration the electrons reach the screen with considerable velocity, and passing through it; impinge on the anode with such force that electrons are set free from its surface; as many as twenty electrons may be emitted for each one arriving. The emitted electrons travel in the direction of the strongest attractive field, that is toward the screen, and an electron current is established from anode to screen. The anode current is then the difference between the rate at which electrons reach it, and the rate at which they leave, while the screen current is the sum of the rate at which electrons arrive from the anode and rate at which electrons arrive from the cathode. The result is that an increase of anode potential causes a fall of anode current and a rise of screen current. When the anode potential approaches equality with that of the screen, the field surrounding the anode exerts a force on the secondary electrons which overcomes the attraction of the screen, so that any secondary electrons emitted are immediately reattracted to the anode. The anode current then rises with an increase of anode potential, until the latter exceeds that of the screen, when the anode current becomes nearly constant and independent of the anode potential.

36. From the load line drawn in fig. 25 it is evident that the valve suffers from certain limitations. Assuming a working anode potential of 120 volts, and grid bias -1.5 volts, the given load line represents a dynamic resistance of 64,000 ohms, a not unlikely figure for the tuned circuit generally used. The distance OA being equal to OB, there will be little or no distortion if the input voltage is not allowed to exceed .5 volt. The power expended in the anode load will then be $\frac{.625 \text{ m.a.} \times 40 \text{ volts.}}{8} = 3.125$ milli-watts. This is the order of the maximum undistorted power

obtainable from this valve and therefore it is unsuitable for use as an "output" or power valve, its use being practically confined to radio-frequency amplification. Further limitations of the S.G. valve, when used as a radio-frequency amplifier for R/T reception, are dealt with in Chapter XII. For this function the tetrode has been largely replaced by the radio-frequency pentode.

CHAPTER VIII.—PARA. 37

The pentode

37. This type of valve has five electrodes, which are termed the cathode, control grid, auxiliary grid, earthed grid and anode, their relative positions being indicated by fig. 27. The auxiliary grid and anode are connected externally to sources of high positive potential, and the earthed grid is connected internally to the cathode. In service valves the auxiliary grid external terminal is mounted on the side of the base. Other terms used for the earthed grid are "suppressor grid" and "anti-secondary."

The design of the pentode arose from the inability of the usual form of screen-grid receiving valve to handle more than a few milli-watts without considerable distortion, which makes it unsuitable for, e.g. the output stage of an amplifier supplying power to a moving coil loud speaker.

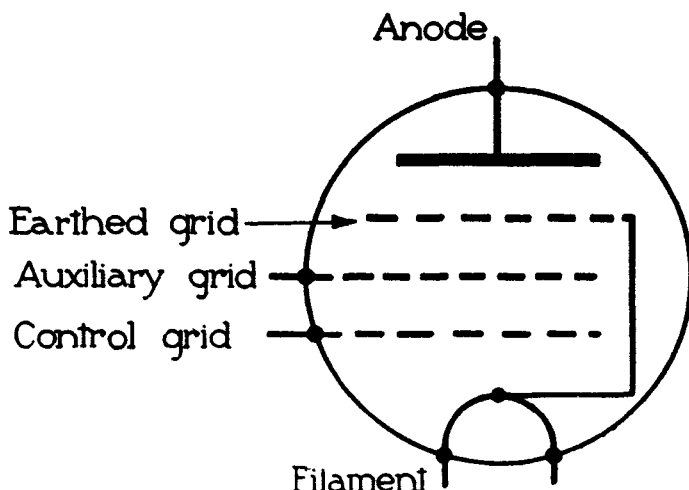


FIG. 27, CHAP. VIII.—Electrodes of pentode valve.

This is due, in part to the limited variation of anode voltage which can be accommodated on the (comparatively) straight portion of its characteristics; a limitation which could be reduced if the region of negative slope were eliminated. The auxiliary grid is retained because as we have seen in the case of the tetrode, an additional electrode with high positive potential has the effect of increasing the anode A.C. resistance and consequently the amplification factor.

When the pentode is designed for low frequency purposes, the auxiliary grid is of comparatively coarse mesh, because the effects of inter-electrode capacitance are not so detrimental as in radio-frequency circuits. The function of the earthed grid is to eliminate the negative slope of the $I_a - V_a$ curve. It will be remembered that this is due to secondary emission from the anode causing an electron current in the direction of the attractive field of the screening grid. In the pentode, the interposition of the earthed grid between the auxiliary grid and the anode screen these secondary electrons from the influence of the auxiliary grid, so that no region of negative slope occurs.

Typical mutual characteristics and $I_a - V_a$ characteristics are given in fig. 28 for comparison with those of the triode and tetrode. The constants of commercial types of pentodes used for power purposes are of the following order :—

Anode A.C. Resistance.	Amplification Factor.	Mutual Conductance.
30,000 ohms.	75	2.5 m.a. per volt.

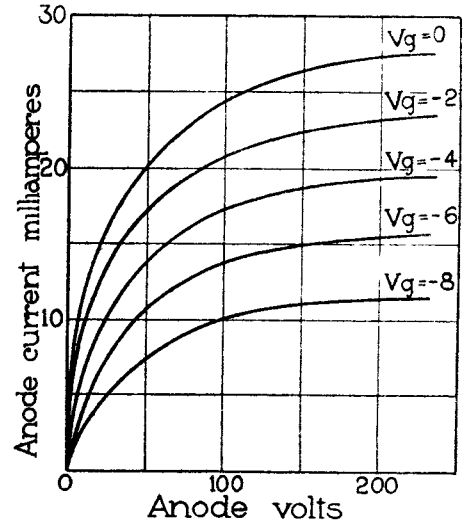
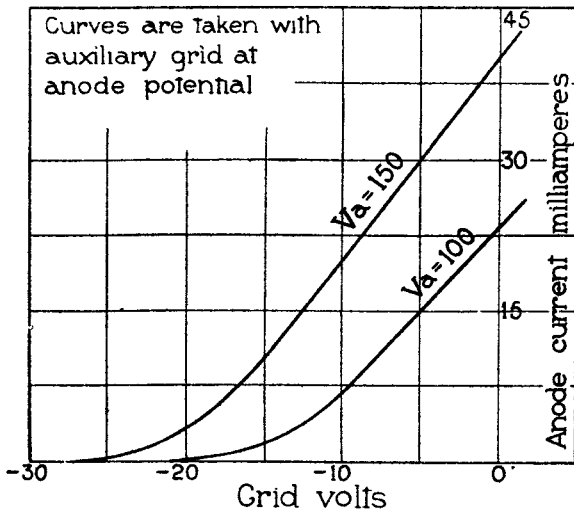


FIG. 28, CHAP. VIII.—Characteristic curves of pentode valve.

Multi-electrode valves

38. Many types of valves have been developed for special purposes, in particular as frequency changers in super-heterodyne receivers. These introduce no new principles, and such as are likely to be found in service radio apparatus will be described with reference to their special function.

CHAPTER IX.—THE RADIO-TELEGRAPHIC TRANSMITTER

THE VALVE OSCILLATOR

Conversion of mechanical energy into oscillations

1. In this chapter it is proposed to deal with the production of oscillatory currents by means of the thermionic valve. By this is meant the conversion of electrical energy, which is supplied by a source of electromotive force, such as a primary battery or a direct current generator, into energy which is still electrical, but which takes the form of oscillations in a circuit possessing inductance and capacitance. In considering the production of continuous electrical oscillations, two useful analogues may be given. First, we have the production of a continuous sound wave by the violin. The sound is a longitudinal vibration of the air which is set in motion by the vibrating violin string. Suppose the frequency to be emitted is 256 cycles per second, corresponding on the scientific scale to the note "middle:C"; it is quite impossible for the performer directly to supply to the string the energy required to cause it to vibrate with undamped amplitude at this rate. To put it crudely, he cannot shake the string continuously at such a high frequency, while merely plucking it and allowing it to vibrate freely will set up a damped vibration, which is not the aim of the performer. The violinist therefore supplies the energy not directly to the string, but to the bow, which is composed of horse hair and is covered with resin. This forms an adhesive surface which repeatedly grips and releases the string as it is moved over the latter, and the continuous supply of energy imparted to the bow is converted into energy in the form of oscillation in the string and the surrounding air. Second, take an ordinary carbon microphone and a telephone receiver and connect them in series with a suitable battery. If the microphone and receiver are placed near each other, it may be found that the telephone receiver commences to howl, apparently spontaneously. If the reader has never performed this experiment, he is advised to do so; if a service "hand press" microphone is used, the type of receiver used for land-line telephone will be found better than a wireless telephone receiver for this purpose, and a six-volt battery will probably be sufficient to cause the effect.

2. This emission of sound by the telephone takes place because sooner or later, some slight noise will occur in the vicinity of the microphone. This noise is a sound wave and consists of successive states of compression and rarefaction in the surrounding atmosphere. These variations of pressure, impinging upon the diaphragm of the microphone, cause variations of resistance of the carbon granule pack, and consequently a variation of current in the electrical circuit. This variation of current flowing round the coils of the telephone receiver produces a movement of the diaphragm, and consequently an emission of sound. This sound travels through the air to the microphone causing a further change in its resistance and a repetition of the foregoing cycle of events will occur, *ad infinitum*. In this instance energy is drained continuously from the battery, and is partly converted into mechanical oscillation of the telephone diaphragm.

The triode oscillator

3. The phenomena associated with the discharge of a condenser in a circuit possessing inductance and resistance is discussed in Chapter VII, and it is there stated that the discharge will be of an oscillatory nature if the resistance of the circuit is less than $2\sqrt{\frac{L}{C}}$. The quantity of electricity stored in the condenser at the end of each successive half-cycle becomes progressively less until all the original energy has been dissipated, and the oscillation so produced is referred to as a damped oscillation. An undamped electrical oscillation will be produced only if arrangements are made to introduce into the circuit, at regular intervals, an amount of energy equal to that dissipated in the interval, and this may be accomplished by means of the triode valve. Fig. 1 shows what is probably the simplest arrangement for the purpose; in this diagram the

CHAPTER IX.—PARA. 4

circuit to be maintained in oscillation consists of a condenser having a capacitance of C farads, an inductance of L henries and a resistance of R ohms which may be supposed to represent the whole of the causes of energy dissipation, while T is a triode valve with its usual sources of H.T. and L.T. supply. The coil L_1 , shown in dotted line, is called the reaction coil, and performs a most important function, but in a preliminary study of the action of the valve, we may consider this coil to be omitted, the grid and filament being directly connected. On closing the switch S an anode current will be established owing to the electric field set up between filament and anode by the H.T. voltage, and this current will flow through the inductance L . The latter must offer opposition to the growth of the current in the form of a counter-E.M.F. the effect of which is to set up a P.D. between the plates of the condenser C , and a resulting displacement current in the dielectric. The anode current tends to a steady value, and the counter-E.M.F. of the inductance becomes smaller and smaller, as the voltage of the condenser rises, and very shortly after the switch is closed, the charge gained by the condenser causes its P.D. to rise above the counter-E.M.F. of the inductance. The condenser will then commence to discharge through the conductive path, setting up a damped oscillation of the nature described in Chapter VII, and upon its cessation,

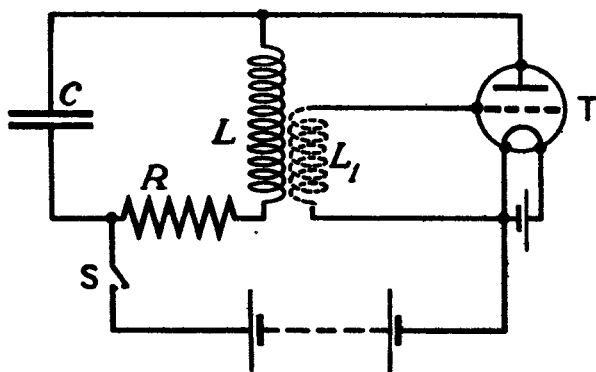


FIG. 1, CHAP. IX.—Simple C.W. oscillator.

the anode current will remain at a constant value. Since there is then no counter-E.M.F. in the inductance no further oscillation will take place. Under the above conditions, the triode merely acts as a series resistance of a few thousand ohms and tends to damp out rather than to maintain the oscillation.

4. Let us consider this oscillation more closely, and suppose that the initial voltage to which the condenser is charged is 100 volts, while the persistency of the ensuing oscillation is 81 per cent. The latter statement implies that if no further energy is imparted to the oscillatory circuit, the amplitude of the condenser P.D. at its second peak will be 90 volts, the polarity of the condenser being reversed. In order to maintain an undamped oscillation an additional charge must be introduced into the condenser during every cycle of oscillation, the magnitude of which will be sufficient to raise the condenser P.D. to its original value despite the wastage of energy which occurs. This energy may be introduced by causing a variation of anode current of such a nature that the resulting counter-E.M.F. is in phase with the counter-E.M.F. caused by the condenser discharge, and the variation of anode current can be achieved by variation of the grid-filament potential. It is for this purpose that the reaction coil L_1 is fitted. The changing flux set up by the oscillatory current through the coil L links with the reaction coil, setting up in it an E.M.F. and therefore a P.D. between grid and filament of the valve, which varies in the same manner as the magnetic flux. The variation of grid-filament potential in turn causes the anode current to vary at the frequency at which the circuit L, C, R , is oscillating; provided that the connections to the coil L_1 are correctly made, the changes of anode current are so timed, or phased, as to augment the charge which is flowing into the condenser at any instant, and the valve will then tend to maintain the circuit in oscillation. Oscillations will however only be maintained if the mutual inductance between the coils L and L_1 exceeds a certain value which depends upon the constants of the

valve as well as upon those of the oscillatory circuit. It has already been stated (Chapter VIII) that if the grid-filament voltage of the triode varies in a sinusoidal manner, i.e. if $v_g = \mathcal{V}_g \sin \omega t$ the valve may be considered to act as a generator having an E.M.F. equal to μv_g or $\mu \mathcal{V}_g \sin \omega t$ volts and an internal resistance of r_a ohms. The equivalent circuit is therefore that of fig. 2 in which the equivalent generator is shown to be supplying the tuned anode circuit L, C, R . As the grid-filament voltage is actually derived from the oscillatory circuit by mutual induction between the anode circuit inductance and the reaction coil, the tuned anode circuit must be in parallel resonance with the supply frequency, and will therefore offer an effective or dynamic resistance of $\frac{\omega^2 L^2}{R}$ or $\frac{L}{CR}$ ohms.

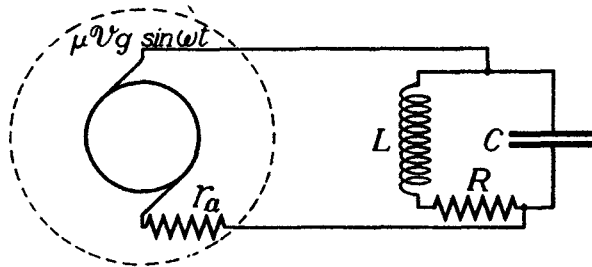


FIG. 2, CHAP. IX.—Equivalent circuit of fig. 1.

Coupling conditions for maintenance of oscillations

5. The building up of an oscillation may now be considered more closely. Assuming that the initial rise of current will cause some feeble oscillation in the circuit L, C, R , let the oscillatory current through the inductance be $i_L = \mathcal{I}_L \sin \omega t$ superimposed upon the steady current, I_a . In this discussion it is only necessary to consider the oscillatory current, and the steady component will therefore be ignored. Remembering that i_L is sinusoidal in form, it may be helpful to omit the factor " $\sin \omega t$ " for the present, and to confine the investigation to amplitudes only. The sequence of events following the initial rise of current may now be enumerated quantitatively:—

(i) The oscillatory current through the inductance L will have a peak value $\mathcal{I}_L$ by hypothesis. This current produces an oscillatory flux, which linking with the coil produces by Faraday's law, an E.M.F.

$$\mathcal{V}_g = \omega M \mathcal{I}_L$$

(ii) This E.M.F. is applied between grid and filament of the valve; the effective voltage in the anode circuit is $\mu \mathcal{V}_g$ and

$$\mu \mathcal{V}_g = \omega M \mu \mathcal{I}_L$$

(iii) The resulting anode current change is of the same form as $\mathcal{V}_g$ that is, sinusoidal, and has the value $\mathcal{I}_a$ where $\mathcal{I}_a = \frac{\mu \mathcal{V}_g}{Z}$; since $Z = r_a + \frac{L}{CR}$

$$\mathcal{I}_a = \frac{\omega M \mu \mathcal{I}_L}{r_a + \frac{L}{CR}}$$

(iv) This oscillatory current acts as a supply current to the rejector circuit L, C, R and will produce in it an oscillatory current $\mathcal{I}'_L$. The relation between supply current $\mathcal{I}_a$ and oscillatory (or circulating) current $\mathcal{I}_o$ in a rejector is given in Chapter V by the equation

$$\mathcal{I}_o = \frac{\omega L}{R} \mathcal{I}_a$$

CHAPTER IX.—PARA. 6

In this particular case $\mathcal{I}_0$ is the new oscillatory current $\mathcal{I}_L$ and the supply current is $\mathcal{I}_a$.

$$\therefore \mathcal{I}'_L = \frac{\omega L}{R} \mathcal{I}_a$$

or

$$\mathcal{I}'_L = \frac{\omega L}{R} \times \frac{\omega M \mathcal{I}_L}{r_a + \frac{L}{CR}}$$

from which it will be seen that the second impulse of oscillatory current $\mathcal{I}'_L$ will be equal to the first one, $\mathcal{I}_L$, if

$$\frac{\omega^2 LM \mu}{R r_a + \frac{L}{C}} = 1.$$

By the relation $\omega^2 = \frac{1}{LC}$ this may be written as

$$\frac{\mu M}{CR r_a + L} = 1.$$

6. If μM is greater than $CR r_a + L$ the second impulse will be greater than the initial one, and as M may be either positive or negative its sign will decide whether the tendency of the

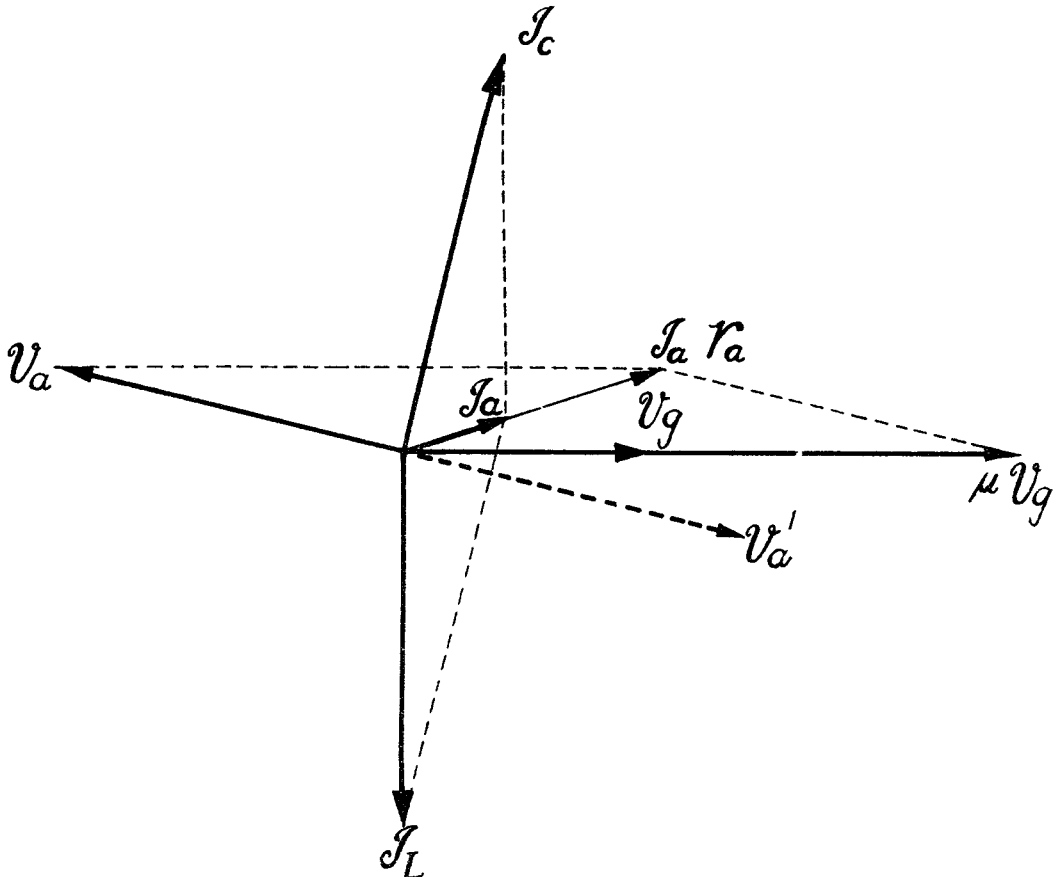


FIG. 3, CHAP. IX.—Vector diagram showing phase relations for maintenance of oscillations.

induced grid-filament voltage v_g is to maintain or suppress the oscillation. Oscillation will only be maintained if the grid-filament voltage leads upon the current in the inductance by an angle approaching 90° , as shown in the vector diagram, fig. 3. Here the datum vector is $\mathcal{I}_L$. Reverting to the use of instantaneous values, since $i_L = \mathcal{I}_L \sin \omega t$ and $\frac{di_L}{dt} = \omega \mathcal{I}_L \cos \omega t$ (Chapter V),

$$\begin{aligned} v_g &= -M \frac{di_L}{dt} \\ &= -M \omega \mathcal{I}_L \cos \omega t. \end{aligned}$$

Therefore v_g lags on i_L by 90° if M is positive, and for the correct phasing conditions M must be negative in sign.

The anode-filament P.D. is shown by the vector $\mathcal{V}_a$ equal and opposite to the vector $\mathcal{V}'_a$ which is the oscillatory P.D. across the tuned anode circuit. Since

$$\mathcal{V}'_a = R i_L + L \frac{di_L}{dt}$$

and

$$\begin{aligned} i_L &= \mathcal{I}_L \sin \omega t \\ \mathcal{V}'_a &= R \mathcal{I}_L \sin \omega t + \omega L \mathcal{I}_L \cos \omega t \\ &= \sqrt{R^2 + \omega^2 L^2} \mathcal{I}_L \sin \left(\omega t + \tan^{-1} \frac{\omega L}{R} \right). \end{aligned}$$

That is, $\mathcal{V}'_a$ leads on $\mathcal{I}_L$ by an angle $\tan^{-1} \frac{\omega L}{R}$ which is nearly 90° because in practice ωL is much larger than R . The anode-filament P.D. $\mathcal{V}_a$ is equal and opposite to this, and is therefore practically 180° out of phase with the grid-filament voltage $\mathcal{V}_g$.

The current $\mathcal{I}_c$ through the capacitive branch of the tuned circuit leads on the P.D. $\mathcal{V}'_a$ by 90° , and the oscillatory component of the anode current, $\mathcal{I}_a$, is the vector sum of $\mathcal{I}_L$ and $\mathcal{I}_c$. It is very nearly in phase with the grid-filament voltage $\mathcal{V}_g$.

Final amplitude reached by the oscillation

7. In the simple circuit shown in fig. 1 the condition for any oscillation once started to increase in amplitude, is that μM must be greater than $CR r_a$. Now although μ , r_a and g_m are commonly referred to as the "constants" of a triode, the ratios which they represent are only constant over the portion of the characteristic curve which is approximately straight. As the anode current decreases, reaching the curved foot of the characteristic, the mutual conductance decreases and the anode A.C. resistance increases. The result of this opposite tendency is that the amplification factor usually remains fairly constant over a very wide range, although the anode A.C. resistance becomes greater for small values of anode current. It may therefore be assumed that in the relation $\mu M > CR r_a + L$ which must be satisfied if the oscillation is to increase in amplitude, all quantities are constant except r_a . If the mutual inductance M is very little greater than that required to make the two sides of the expression equal, the oscillation will increase in amplitude until the variations of anode current extend to a portion of the characteristic over which the average value of r_a is such that $CR r_a + L$ becomes equal to μM , and the oscillation is maintained at this amplitude.

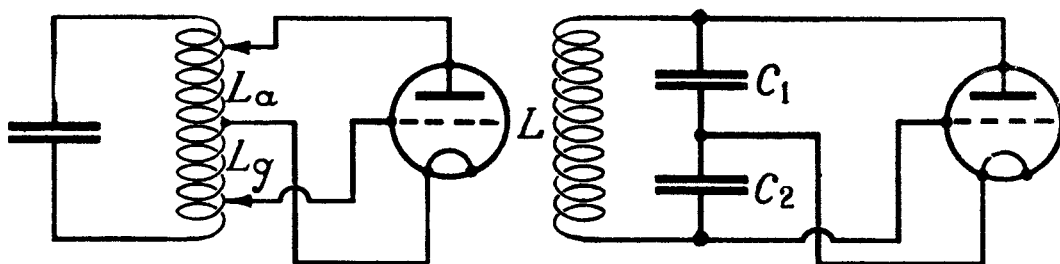
Valve oscillator circuits

8. Although it is necessary for some form of coupling to exist between the grid and anode circuits in order to maintain the anode current in oscillation it is not necessary that this coupling shall be due to mutual induction. Any form of inductive or capacitive coupling can be utilised provided the connections are so made that the correct phase relationship is obtained between grid and anode oscillatory voltages. Three common arrangements are shown in fig. 4 in which the method of supplying the H.T. voltage to the anode has been disregarded for the sake of simplicity.

CHAPTER IX.—PARA. 8

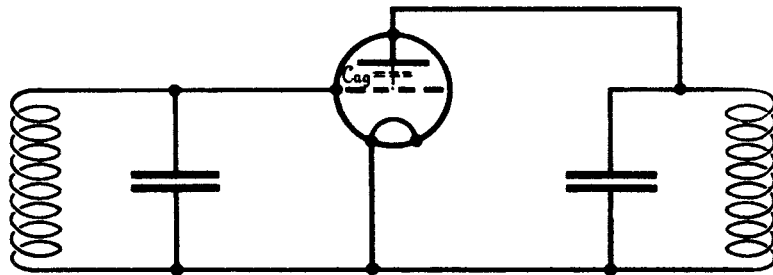
The Hartley oscillator.—The scheme of connections shown in fig. 4a makes use of combined mutual and auto-inductive coupling and is called the Hartley circuit. This is frequently used for oscillators operating upon the higher radio frequencies of the order of from 3 to 30 megacycles per second, interchangeable coils being of course necessary to cover the whole band. Its practical advantage lies in the ease with which sufficient coupling may be attained even with coils of only two or three turns. If the coil is wound continuously in a spiral the oscillatory anode-filament and grid-filament voltages are essentially in approximate anti-phase because the filament is connected to some point between the grid and anode, and the current in the winding L_a lags by 90 degrees on the grid-filament voltage.

The Colpitts oscillator.—In this circuit, fig. 4b, the oscillation is maintained by auto-capacitive coupling. This circuit is used in some service transmitters particularly for the lower frequencies,



(a) Hartley circuit

(b) Colpitts circuit



(c) Tuned anode - Tuned grid circuit

FIG. 4, CHAP. IX.—Valve oscillator circuits.

i.e. of the order of 200 k.c/s. The oscillatory circuit itself consists of the inductance L and the two condensers C_1 and C_2 in series. Suppose that on switching on some slight oscillation takes place in the circuit L , C_1 , C_2 . The oscillatory current in the capacitive branch sets up a corresponding P.D. between the plates of the condenser C_2 and as these are connected between grid and filament this voltage causes a further change of anode current, which is so phased as to reinforce the original oscillation. The condensers C_1 and C_2 together form what is to all intents and purposes a potentiometer, a portion of the anode-filament P.D. being tapped off to supply the grid excitation, and the circuit is inherently self-oscillatory provided that the ratio $\frac{C_2}{C_1}$ does not exceed the amplification factor of the valve. In a practical transmitter the capacitance C_1 is usually that of the aerial, while C_2 has a fixed value depending upon the type of valve and the type of aerial it is required to use. The transmitter may then be made to cover a wide frequency

range by using interchangeable tuning inductances only. A transmitter using mutual inductive coupling for the same frequency range would require two sets of interchangeable coils and a larger number of connecting plugs and sockets, which tend to cause a loss of efficiency and power output owing to dirt and corrosion.

The tuned-anode/tuned-grid circuit.—This is shown in fig. 4c. Anode and grid circuits possess both inductance and capacitance and are brought into resonance or very nearly into resonance with each other. In such circumstances the inherent inter-electrode capacitance between grid and anode (and the connecting leads thereto) is sufficient to transfer energy from one circuit to the other, the resulting grid-filament P.D. causing a variation of anode current of such a phase as to maintain the anode circuit in oscillation. This circuit is most suitable for high frequency transmission of the order of 3 to 15 megacycles per second. To cover such a wide frequency band several sets of coils are required, generally in pairs, although sometimes a single grid coil may cover the range of two anode coils. The anode circuit capacitance may, of course, be that of the aerial.

Series and shunt feed

9. In the oscillator circuit given in fig. 1 the H.T. supply is connected between one end of the tuning inductance and to the filament. The tuning coil then carries both the oscillatory current and the steady anode current of the valve, and the arrangement is known as the series feed system. An alternative method of feeding is the parallel feed system, fig. 5, in which the

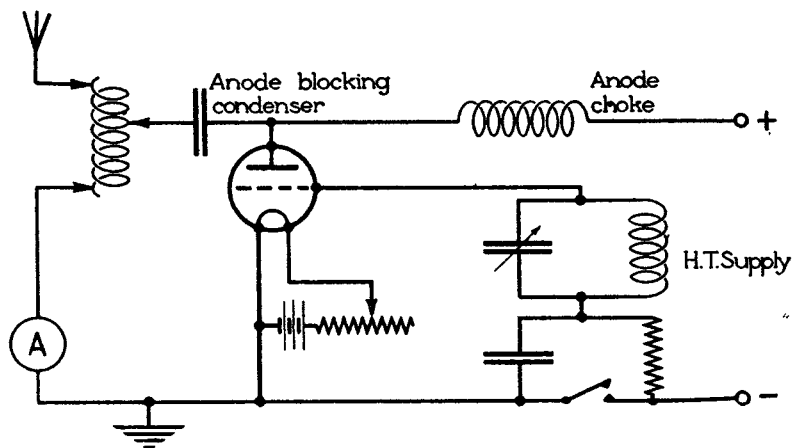


FIG. 5, CHAP. IX.—Parallel feed to valve oscillator.

H.T. supply is fed through a choking coil, the inductance of which is large compared to the inductance of the tuning coil. The H.T. supply via this anode choke, as it is termed, is connected directly to anode and filament of the valve, and the tuning inductance is normally also connected between the same points, but a condenser called the anode blocking condenser is interposed in order that the H.T. supply shall not be short-circuited by the tuning inductance. There is no essential difference in the action of series and parallel feed systems, but the latter has the advantage that the aerial tuning inductance is insulated from the source of H.T. supply. The anode choke must be designed with great care in order to ensure that its natural frequency does not fall within the frequency band to be covered by the transmitter. If this does occur, heavy circulating currents may flow in the coil and cause a breakdown, or alternatively the damping losses may be so great that the valve fails to maintain the oscillation. Interchangeable anode chokes for different frequency bands form a possible solution but only at the expense of an increase in cost, space, weight and rapidity of frequency changing.

CHAPTER IX.—PARA. 10

Evolution of a C.W. transmitter

10. The circuits shown in figs. 1 and 4 are generally referred to as C.W. oscillators to distinguish them from actual transmitting circuits in which many additions to the simple oscillator circuits are generally fitted. The circuit of a typical transmitter in which many of these refinements have been incorporated is given in fig. 6, in which the devices enumerated in the following sub-paragraphs are indicated by the corresponding arabic numeral.

(i) The oscillations are generated either in an "aerial circuit" tuned to the desired frequency, which is called direct aerial excitation, or are caused to set up oscillations in it by additional apparatus. The latter methods of aerial excitation are now becoming universal, and are dealt with later. For the present attention will be confined to the direct form of aerial excitation, and the oscillations will be considered to be generated in the aerial circuit, which is an open oscillator and consequently a good radiator.

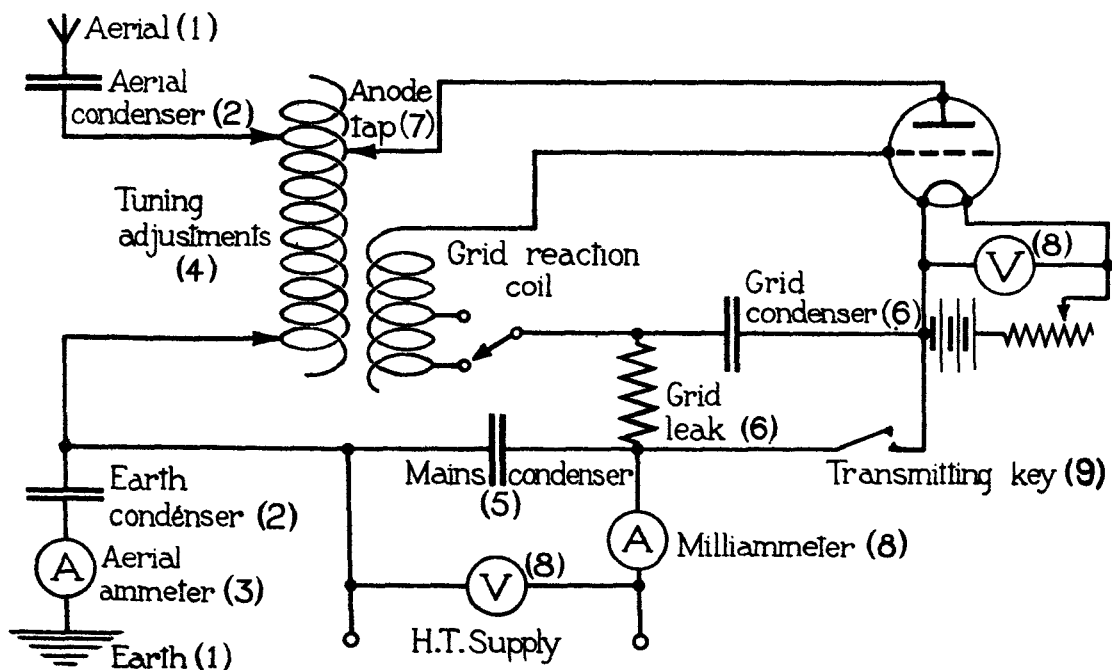


FIG. 6, CHAP. IX.—C.W. transmitter, direct aerial excitation.

(ii) With the method of H.T. supply shown in fig. 1, if the upper and lower plates of the condenser *C* are replaced by an aerial and earth connection respectively, a possible source of failure and danger to personnel is introduced, for both the aerial wire and the earth connection are now at a high steady potential with respect to the filament of the valve. A person standing on the earth and touching the filament battery, for example, would "short-circuit" the source of H.T. supply. To avoid this, large condensers are placed in series with the aerial and earth connections, the order of capacitance being $\cdot 01 \mu F$. Such large condensers cause practically no decrease in the total effective capacitance of the aerial-earth system.

(iii) An aerial ammeter is provided in order to show when oscillations are being produced, and to indicate the magnitude of the oscillatory current. Operators frequently under-estimate the importance of the aerial ammeter; it must always be remembered that if a transmitter is switched on and the key pressed, the existence of a fault in the oscillatory circuit is generally indicated by the absence of "reading" on this meter. If oscillations do not commence when

power is supplied to the valve, serious heating of the latter may occur. Especially in conditions where the valve is out of sight of the operator as in some aircraft sets, the aerial ammeter must be regarded as an essential part of the equipment and not as a refinement which could be dispensed with.

(iv) Means are provided for variation of frequency of the generated oscillations. This is usually achieved by varying the amount of inductance in the aerial circuit, the exact method depending upon the requirements of the particular transmitter.

(v) A mains condenser is provided and is connected across the terminals of the source of H.T. supply; its value should be large and its insulation resistance high. The function of this condenser is to confine the oscillatory variations of anode current to the transmitter itself. If the H.T. source is a D.C. generator, its armature windings will be of high impedance, and will seriously reduce if not entirely prevent the variation of anode current which is necessary to maintain the oscillation. The condenser forms a low impedance path in parallel with the generator windings, allowing these variations to reach the full value permitted by the valve characteristic and dynamic resistance of the oscillatory circuit.

(vi) Suitable arrangements are made for the provision of a considerable grid bias voltage. The necessity for this will be considered later.

(vii) The anode of the valve is connected to some suitable point on the aerial tuning inductance, which is called the anode tapping point. This requirement also will be dealt with in due course.

(viii) For the purpose of adjusting the power supplies to the transmitter, suitable ammeters and voltmeters are provided. If weight and space permit, these are (a) a voltmeter showing the voltage applied to the filament. (b) a milliammeter showing the average anode current. (c) a voltmeter showing the voltage of the H.T. supply.

(ix) Means must be provided for interrupting the wave by a morse key for telegraphic transmission. The method of "keying" the transmitter usually adopted in aircraft sets using direct aerial excitation is to interrupt both the negative H.T. lead and the grid circuit. For high power or high-speed (automatic) transmission the "keying" arrangements may be more complicated e.g. as in figs. 29 and 42.

Production of I.C.W. waves

11. The radiation produced by a C.W. transmitter is of constant amplitude throughout the period during which the transmitting key is pressed, except for a very short period during which the oscillation is building up to its maximum amplitude, and a similar period when the key is

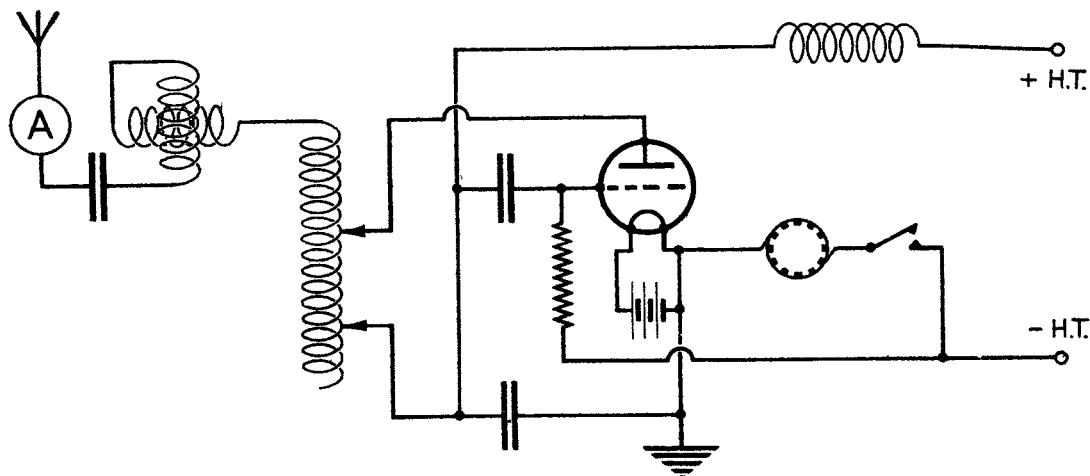


FIG. 7, CHAP. IX.—I.C.W. transmitter.

CHAPTER IX.—PARA. 12

raised, for the aerial will then continue to oscillate with rapidly diminishing amplitude until the energy stored at the moment of raising the key is totally dissipated. If it is desired to radiate I.C.W. waves, the continuous wave must be interrupted at some rate which lies in the audio-frequency range, in practice from 200 to 2,000 times per second. The simplest method, which is also as efficient as any other, is to arrange a rotary interruptor in series with the transmitting key, fig. 7. A typical form of interruptor consists of a pair of brass discs which are mounted side by side upon the shaft of a motor and are in electrical connection, but insulated from the shaft. One of the discs is wholly of metal and is known as the slip ring, while the interruptor disc proper has a number of insets of insulating material (usually mica or fibre) let into its periphery. Brushes of springy phosphor bronze bear upon the edge of each disc. When the transmitting key is pressed and the discs are in rotation, the current in the key circuit is interrupted at a rate depending upon the speed of rotation and the number of conducting segments on the edge of the interruptor disc, i.e. if the speed is 2,400 r.p.m. or 40 r.p.s. and the disc has 10

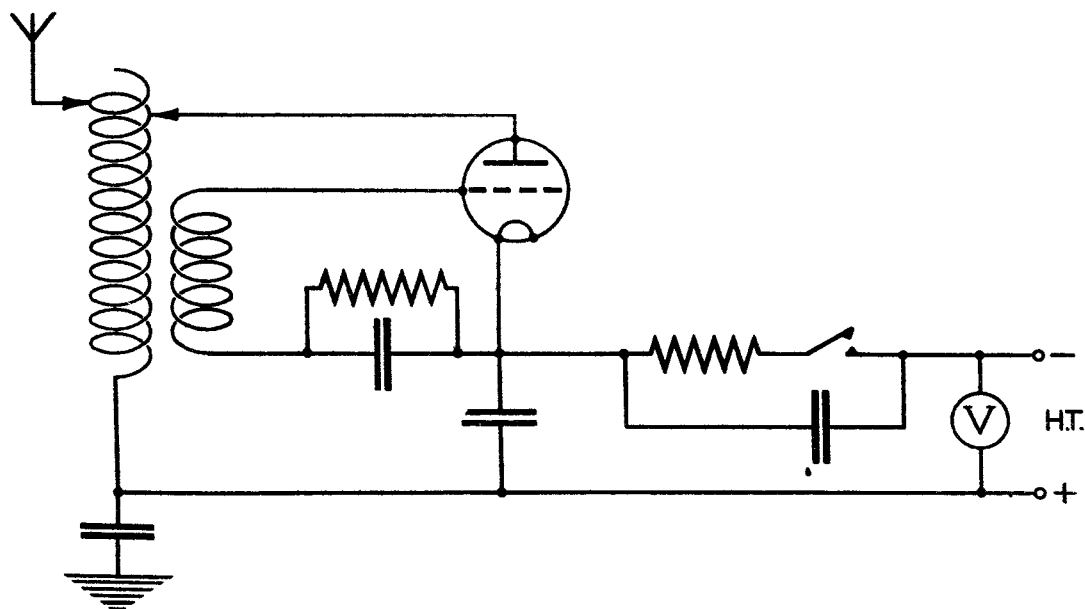


FIG. 8, CHAP. IX.—Simple key click eliminator.

conducting segments, the number of interruptions will be 400 per second, and the resulting transmission will consist of interrupted continuous waves having this group frequency. This method has the advantage that the interruptor may be mounted on the shaft of the generator supplying the anode voltage for the transmitter, which is particularly convenient in aircraft.

Key clicks

12. If a high frequency receiver is situated in the vicinity of a medium or low frequency transmitter or vice versa, it is often found that severe interference to reception is caused by the operation of the transmitter in spite of the large difference in frequency between the radiated wave and that to which the receiving aerial is tuned. This interference has been traced to two causes, (i) radiation of harmonics by the transmitter, (ii) the shock excitation of the receiving aerial. The former effect can only be eliminated by reduction of the harmonic content of the radiated wave, one possible remedy being an indirect coupling between the aerial and the oscillatory circuit of the transmitter. The latter effect, which is generally referred to as key click interference, is produced as follows. When the transmitter is "keyed" by the normal

method, the full H.T. voltage is immediately applied between the anode and filament of the transmitting valve and the resulting oscillation reaches its maximum amplitude in a very short space of time, equivalent to say 20 to 30 cycles. The field strength in the vicinity of the receiver increases in magnitude in the same way, and causes a rapidly increasing induced E.M.F. in the receiving aerial, which is set into oscillation at its natural frequency no matter how remote this may be from that of the inducing electric field. The remedy

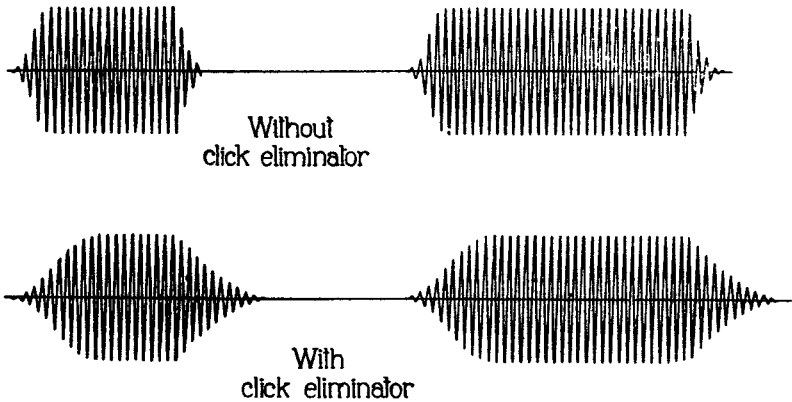


FIG. 9, CHAP. IX.—Growth and decay of oscillations.

for this state of affairs is to ensure that the oscillatory current in the transmitting aerial reaches its final amplitude very slowly, and this may be achieved by either of two methods. In the simplest form, a resistance of the order of 500 ohms is inserted in series with the H.T. supply to the generator, on the generator side of the mains condenser as shown in fig. 8. On closing the transmitting key, the voltage applied to the anode-filament path of the valve is that of the mains

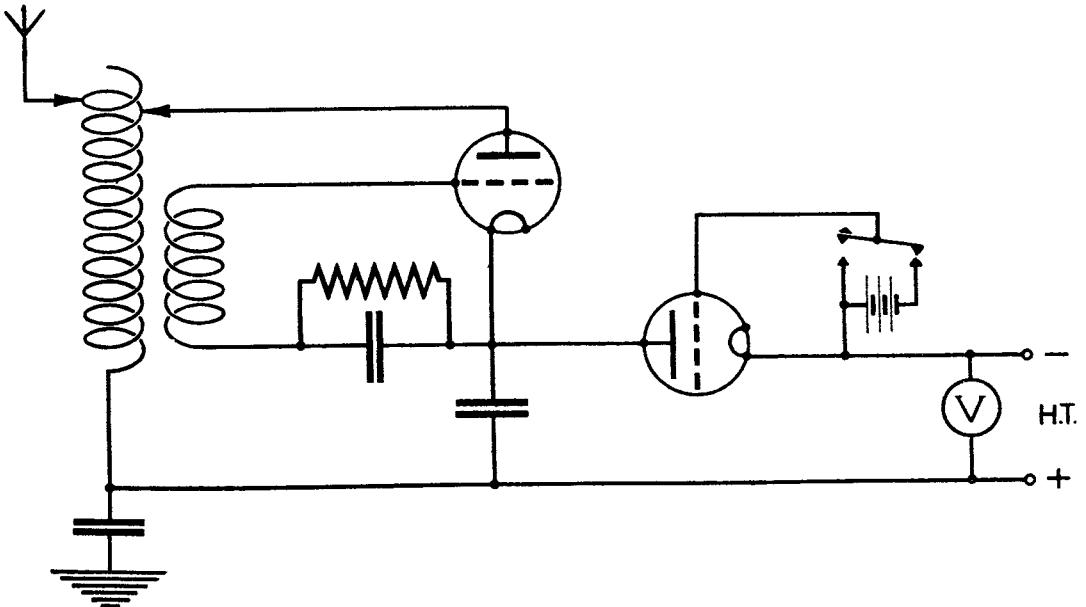


FIG. 10, CHAP. IX.—Key click eliminator for high power transmitter.

CHAPTER IX.—PARAS. 13-14

condenser, and the latter is initially not equal to that of the H.T. supply, owing to the presence of the resistance, although it rises to this value after a certain interval as explained in Chapter I. When the transmitting key is raised, the mains condenser is charged to the supply voltage, and continues to supply the anode of the valve, the voltage gradually falling in a manner complementary to that in which it increases when the key is pressed. The amplitude of the oscillatory current in the aerial therefore increases and decreases slowly, and shock excitation of neighbouring receiving aerials is avoided. A condenser is shunted across both key and resistance to reduce sparking at the key contacts. The effect of the key click eliminator upon the rate of growth and decay of the oscillation is shown in fig. 9. It must be noted that this remedy necessitates a slight change in the location of the transmitting key, in order to ensure that the mains condenser becomes discharged at the end of each morse element.

13. A device which is more suitable for installation in high power transmitters is shown in fig. 10. Here the resistance element of the key click eliminator is the anode-filament path of an auxiliary thermionic valve, and the transmitting key is inserted in its grid circuit. When the key is raised the grid of the auxiliary valve is given a considerable negative potential with respect to the filament, and the resistance of its anode-filament path becomes infinitely large, so that no current can flow in the anode circuit of the transmitting valve. On pressing the key, the grid of the auxiliary valve is given zero potential with respect to its filament, and the resistance of its anode filament path falls to some value of the order of a thousand ohms, allowing the mains condenser to charge and oscillations to build up as in the circuit previously described. The advantage of this method is that the transmitting key is not required to break the main anode circuit, and sparking at the key is negligible; it is therefore better adapted for high power and high speed signalling than its prototype. The principal disadvantage of this device is the necessity for high insulation of the filament supply for the auxiliary valve.

Tuning a transmitter

14. This term is used in the service to signify the whole procedure of preparing a transmitter for operation on a given frequency. The exact details of course vary with different transmitters, but the following sequence of operations is usually necessary.

(i) Examine all external connections. This precaution should never be omitted unless the transmitter has been observed in operation immediately before retuning. After completing a minor repair, it is easy to omit to replace a single lead, which may fall on to another and cause a short-circuit.

(ii) Switch on the L.T. supply to the filament, and adjust the filament voltage or current to the correct value for the type of valve in use.

(iii) Switch on the H.T. supply to the transmitter, adjusting its value to not more than two thirds the maximum permissible voltage for the particular transmitter.

(iv) Having verified that H.T. and L.T. supplies are correct switch them off; switches are generally fitted in such a manner that this can be done without prejudice to the adjustments made in (iii).

(v) Adjust the aerial tapping point on the aerial tuning inductance to a value suitable for the desired frequency. With any given type of transmitter experience is the only guide to this; for this reason many service transmitters are provided with tables of approximate adjustments which should be consulted.

(vi) Set the anode tap, if separately adjustable, to a point not nearer to the tuning tap than half way between the latter point and the earth-potential end of the tuning coil.

(vii) Switch on H.T. and L.T. supply, press the transmitting key and observe (a) the aerial ammeter, (b) the anode current milliammeter; (a) should indicate that oscillatory current is flowing in the aerial circuit, (b) that anode current is flowing; care should be taken that this is not excessive. The normal anode current for the transmitter is usually given in the tables previously mentioned. In their absence it should be remembered that overheating of the valve will be avoided, even if oscillations are absent, provided that the power input to the transmitter

is less than the permissible dissipation of the valve, e.g. if the transmitter has only a single V.T. 5 B valve, rated at 250 watts, and the H.T. voltage is adjusted to 2,000 volts, an anode current not exceeding 125 milli-amperes is permissible.

(viii) If oscillations do not commence, try a change of anode tap. In transmitters with reversible connections to the grid circuit, the appropriate leads should be interchanged. When neither of these adjustments produce the desired effect, some fault probably exists and the steps to be taken are usually detailed in the appropriate Chapter of A.P.1186 (Signal Manual Part IV).

(ix) When oscillations are obtained, the key should be held down and the frequency measured with an appropriate wavemeter; adjustment of the aerial tuning inductance is then performed until the frequency is very near to that desired. There is no necessity to make this adjustment to the limit of possible accuracy, if the anode tap is separately adjustable.

(x) Adjust the anode tap until maximum output (i.e. aerial amperes) is obtained. The anode current will then usually have the minimum value. If this is not so, adjust for the greatest ratio of oscillatory to anode currents, provided that the maximum permissible anode current is not exceeded.

(xi) Readjust the aerial tuning inductance to the frequency desired, with the utmost precision to which the wavemeter can be observed.

(xii) In carrying out these operations two precautions should always be observed. First, the key should never be held down longer than necessary. The practice of locking the key while the tuning chart is consulted is liable to cause overheating of the transmitter components and the H.T. generator. Secondly, remember to switch off the H.T. supply to transmitter before touching any component other than the transmitting key. In modern H.T. transmitters the dangerous portions are usually screened in such a manner that access is impossible without interrupting the H.T. supply.

Adjustment of transmitter to desired frequency

15. This adjustment is performed with the aid of an instrument known as a wavemeter or frequency meter. Many different types of wavemeter are in use in the service, and an account of the principles involved in their design and operation will be found in Chapter XIII. For the present, the simplest form will be briefly described. This consists of a calibrated closed oscillatory circuit, incorporated with which is some device which serves to indicate the presence of an oscillatory current in the circuit. One such device is a miniature 2-volt lamp, which will glow with full

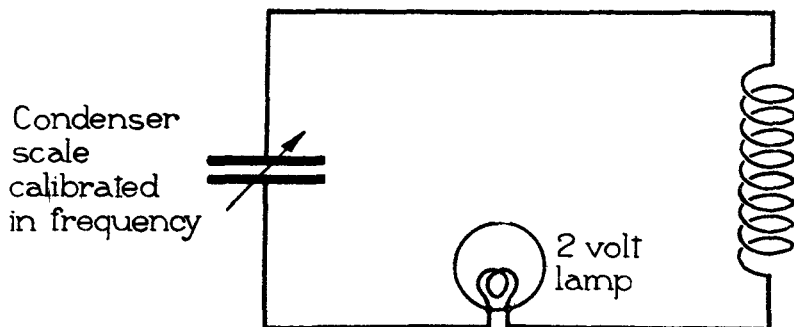


FIG. 11, CHAP. IX.—Simple wavemeter.

brilliance if a current of about 100 milliamperes is passed through it (fig. 11). When the transmitter is in operation, the wavemeter is so held that the magnetic field of the transmitting inductance links with the inductance of the wavemeter circuit, and an oscillatory E.M.F. is induced in the latter. The resulting current in this circuit will reach a maximum if the wavemeter circuit is in resonance with that of the transmitter, and this is shown by the maximum brilliancy of the lamp. The frequency to which the meter is adjusted is then read from its calibrated scale, and this is identical with the transmitter frequency.

EFFICIENCY AND OUTPUT

Oscillatory power in sinusoidal oscillator

16. In order to maintain oscillations in the simple circuit of fig. 1, a portion of the power taken from the H.T. supply must be transferred to the oscillatory circuit. A simple investigation will show how great this quantity must be. The total anode current at any instant will be $I_a + \mathcal{I}_a \sin \omega t$ while the anode-filament P.D. at the same instant will be $E_a - \mathcal{V}_a \sin \omega t$ because $\mathcal{I}_a$ is 180° out of phase with $\mathcal{V}_a$. The power dissipated in the form of heat in the valve itself will be the product of the anode current and the anode voltage and is equal to $(I_a + \mathcal{I}_a \sin \omega t)(E_a - \mathcal{V}_a \sin \omega t) = P_L$.

$$\begin{aligned} P_L &= I_a E_a - I_a \mathcal{V}_a \sin \omega t + \mathcal{I}_a E_a \sin \omega t - \mathcal{I}_a \mathcal{V}_a \sin^2 \omega t \\ &= I_a E_a - I_a \mathcal{V}_a \sin \omega t + \mathcal{I}_a E_a \sin \omega t - \frac{\mathcal{I}_a \mathcal{V}_a}{2} + \frac{\mathcal{I}_a \mathcal{V}_a}{2} \cos 2 \omega t. \end{aligned}$$

Of the above expression, only two terms have average values over a number of complete cycles, viz. $I_a E_a$ and $\frac{\mathcal{I}_a \mathcal{V}_a}{2}$. The power expended in heating the anode is therefore $I_a E_a - \frac{\mathcal{I}_a \mathcal{V}_a}{2}$. Now

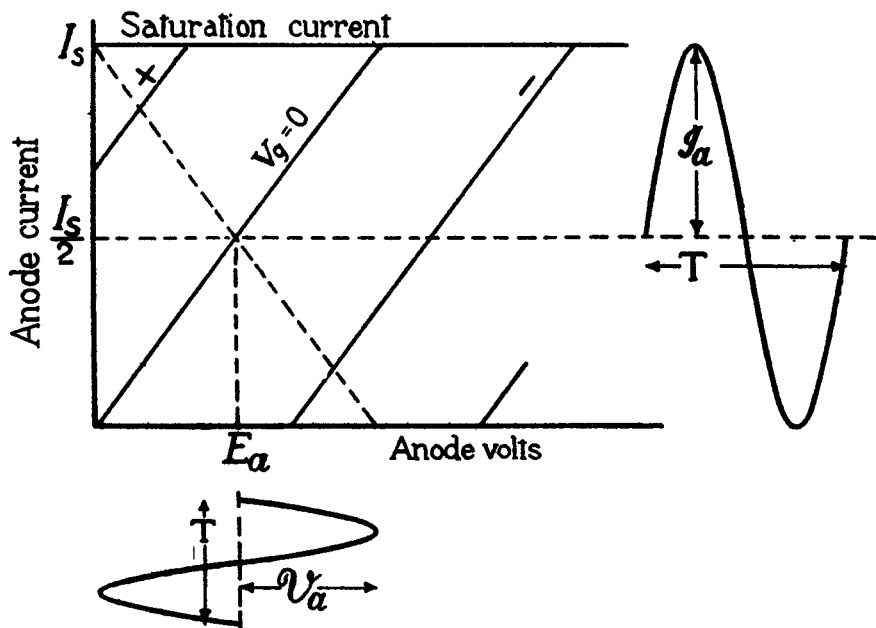
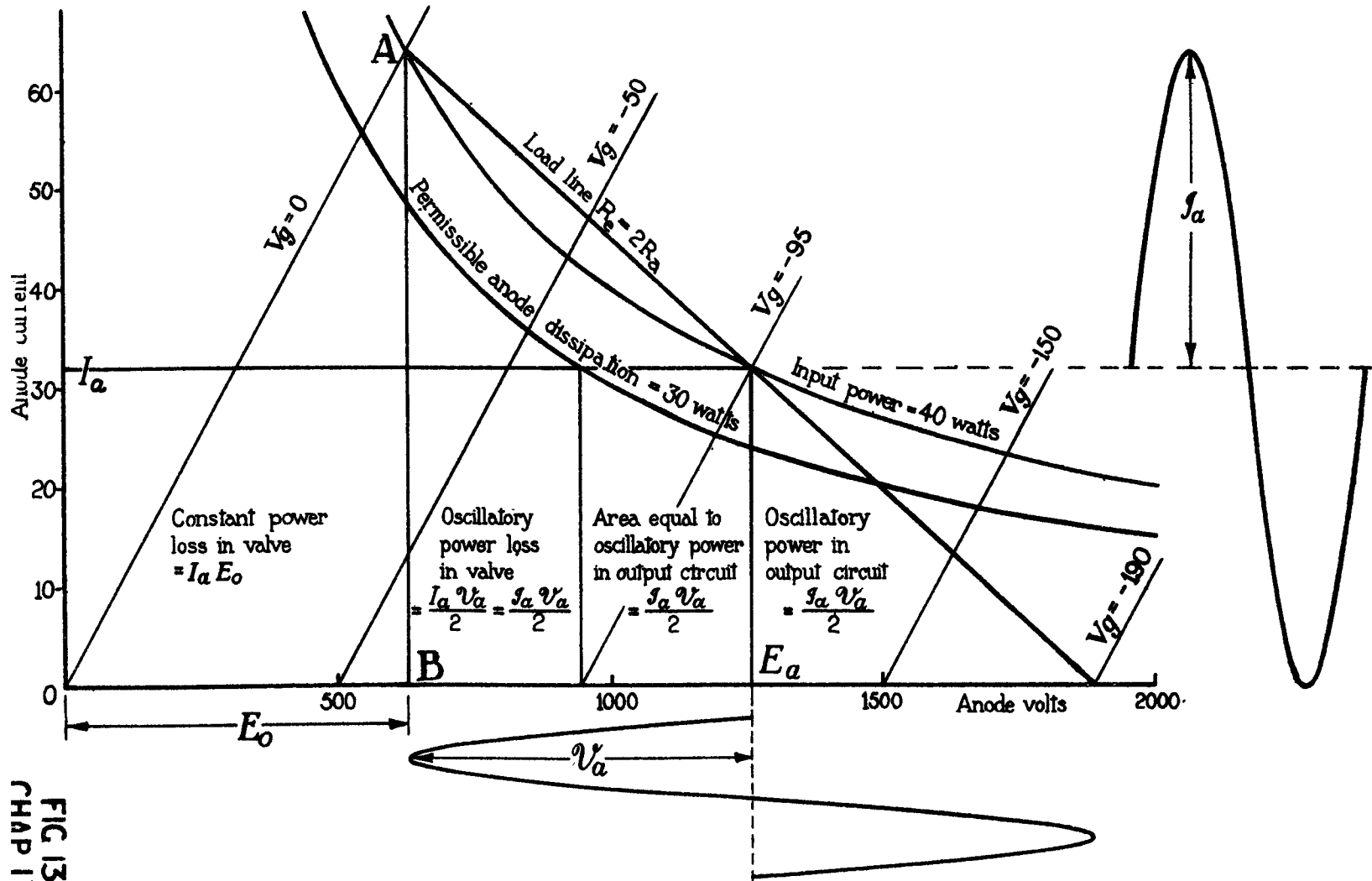


FIG. 12, CHAP. IX.—Ideal $I_a - V_a$ characteristics, showing conditions giving theoretical efficiency of 50 per cent.

$I_a E_a$ is the power supplied to the circuit by the H.T. source, and the difference between $I_a E_a$ and $I_a E_a - \frac{\mathcal{I}_a \mathcal{V}_a}{2}$, namely $\frac{\mathcal{I}_a \mathcal{V}_a}{2}$, is the power which is converted into oscillatory form. If this is to be a maximum, the amplitudes of the oscillatory current $\mathcal{I}_a$ and oscillatory voltage $\mathcal{V}_a$ must be as large as possible. The maximum possible variation of anode current is from zero to saturation value. Suppose that the $I_a - V_a$ characteristic curves of the valve are straight lines between zero and saturation value I_s , the zero grid volts curve passing through the origin



SINUSOIDAL OPERATION WITHOUT GRID CURRENT

(fig. 12). This assumption is frequently made for theoretical purposes and the curves are referred to as "ideal characteristics." Provided that the mean anode current is one-half the saturation value, $\frac{I_s}{2}$, the anode current can execute a variation between zero and saturation value, and $\mathcal{I}_a$ will have the value $\frac{I_s}{2}$. The maximum variation of anode-filament P.D. is from zero to $2E_a$, which is only possible if the amplitude $\mathcal{V}_a$ of the oscillatory component is equal to the steady voltage E_a . The maximum power output is therefore $\frac{\mathcal{V}_a \mathcal{I}_a}{2} = \frac{E_a I_s}{4}$, while the power input is $E_a I_a = \frac{E_a I_s}{2}$. The ratio of power output to power input, i.e. the efficiency of the valve as a power converter, is therefore 50 per cent. A study of fig. 12 will show that there is only one value of external or load resistance $R_e = \frac{L}{CR}$ which will allow this output and efficiency to be obtained under the given conditions. The load line must pass through two points, viz. saturation current when the anode-filament P.D. falls to zero, and one-half saturation value when the anode-filament P.D. is E_a . On drawing the load line through these points, it is obvious that its slope is equal to the slope of the valve characteristics i.e. the dynamic resistance of the load must be equal to the anode A.C. resistance of the valve.

Efficiency and output without grid current

17. It must now be made clear that the above conditions are not possible in practice. The postulation of straight-line characteristics is justifiable because a slight departure from linearity does not affect the output but only the wave-form. The serious error in the above reasoning is the assumption that the amplitude of the oscillatory component of anode-filament P.D. can be made equal to the H.T. supply voltage E_a . As maximum positive grid potential and minimum anode-filament P.D. occur practically simultaneously owing to the approximately anti-phase relationship between $\mathcal{V}_a$ and $\mathcal{V}_g$ (fig. 3) it is essential that the anode potential shall never fall below the most positive value of grid potential, otherwise excessive grid current will flow. The valve will then cease to generate oscillations at the frequency of the tuned circuit, although possibly a spurious or parasitic oscillation may be set up at some other frequency.

No anxiety will arise on this account if steps are taken to ensure that the grid potential can never attain any positive value. Let us therefore investigate the production of oscillations with sinusoidal variation of anode current and voltage; but with sufficient negative grid bias to ensure that no grid current will flow during any portion of the cycle. We may assume that the valve possesses ideal characteristics as before, and that the maximum power dissipation of the valve must not be exceeded; this will generally preclude the possibility of allowing the anode current to rise to saturation value. The characteristics shown in fig. 13 closely resemble those of a small transmitting valve, the permissible dissipation being 30 watts. The load line lies across the characteristics with one end terminating upon the curve $V_g = 0$, because it has been stipulated that the grid voltage variation must not extend into the region of positive grid voltage. The projection of this point upon the voltage axis, as shown by the line A B, gives the minimum allowable anode-filament P.D., which will be denoted by E_o . The H.T. supply voltage will be $E_a = E_o + \mathcal{V}_a$, the mean anode current I_a and the total variation of anode current from zero to $2I_a$. The amplitude of the oscillatory anode current, $\mathcal{I}_a$, will therefore be equal to the mean anode current I_a ; the output power will be $\frac{\mathcal{I}_a \mathcal{V}_a}{2}$ and the input power $I_a E_a$ or $\mathcal{I}_a (E_o + \mathcal{V}_a)$. Without further calculation therefore we may conclude that the efficiency η under these conditions is $\frac{\mathcal{I}_a \mathcal{V}_a}{2} \times \frac{1}{\mathcal{I}_a (E_o + \mathcal{V}_a)} = \frac{\mathcal{V}_a}{2 (E_o + \mathcal{V}_a)}$.

CHAPTER IX.—PARA. 18

The value of E_o in terms of $\mathcal{J}_a$ may be found from the following considerations. In fig. 13, $A B = \delta I_a$ and $E_o = \delta V_a$, for the curve $V_g = 0$. Hence $\frac{E_o}{A B} = r_a$ and since $A B = 2 \mathcal{J}_a$, $E_o = 2 \mathcal{J}_a r_a$.

$$\therefore \eta = \frac{\mathcal{V}_a}{2(2\mathcal{J}_a r_a + \mathcal{V}_a)} = \frac{1}{2\left(2\frac{\mathcal{J}_a}{\mathcal{V}_a} r_a + 1\right)}.$$

As $\frac{\mathcal{V}_a}{\mathcal{J}_a}$ is equal to the load resistance R_e ,

$$\eta = \frac{R_e}{2(2r_a + R_e)}.$$

18. From this it appears that for a given valve the efficiency depends upon the load resistance, and approaches fifty per cent. only if R_e is infinitely large compared to r_a . The power output under these conditions would be extremely small, because a high load resistance entails a small oscillatory current and a corresponding decrease of oscillatory grid voltage. To find the maximum output, the maximum permissible grid excitation must be obtained. As the oscillatory output power is equal to the product of the efficiency and input power, the output is

$$P_o = \frac{R_e}{2(R_e + 2r_a)} \times E_a I_a = \frac{R_e E_a \mathcal{J}_a}{2(R_e + 2r_a)}.$$

Substituting the value of $\mathcal{J}_a$, namely $\frac{\mu \mathcal{V}_g}{r_a + R_e}$

$$P_o = \frac{R_e E_a}{2(R_e + 2r_a)} \times \frac{\mu \mathcal{V}_g}{r_a + R_e}.$$

Another expression for P_o is obtained by noting that it is equal to $\frac{R_e \mathcal{J}_a^2}{2}$ or $\frac{R_e}{2} \times \frac{\mu^2 \mathcal{V}_g^2}{(r_a + R_e)^2}$.

Equating the two expressions for P_o ,

$$\frac{R_e E_a}{2(R_e + 2r_a)} \times \frac{\mu \mathcal{V}_g}{r_a + R_e} = \frac{R_e}{2} \times \frac{\mu^2 \mathcal{V}_g^2}{(r_a + R_e)^2}$$

which simplifies readily to

$$\frac{\mu \mathcal{V}_g}{r_a + R_e} = \frac{E_a}{R_e + 2r_a}$$

and the maximum permissible grid excitation is therefore

$$\mathcal{V}_g = \frac{E_a}{\mu} \times \frac{R_e + r_a}{R_e + 2r_a}.$$

The grid bias will have the same value, locating the mean operating point on the load line in such a position that a sinusoidal input to the grid reaches the limits of zero grid voltage on the positive half-cycle and zero anode current on the negative half-cycle. The power output is then a maximum and its magnitude in terms of r_a , R_e and the anode supply voltage E_a can now be determined.

As

$$\mu \mathcal{V}_g = \frac{E_a (R_e + r_a)}{R_e + 2r_a}$$

while

$$\mathcal{J}_a = \frac{\mu \mathcal{V}_g}{R_e + r_a}$$

$$\begin{aligned} P_o &= \frac{\mathcal{J}_a^2 R_e}{2} = \frac{E_a^2 \left(\frac{R_e + r_a}{R_e + 2r_a} \right)^2}{(R_e + r_a)^2} \times \frac{R_e}{2} \\ &= \frac{E_a^2}{2} \times \frac{R_e}{(R_e + 2r_a)^2}. \end{aligned}$$

It can be proved that this is a maximum when $R_c = 2r_a$ and therefore the maximum output is

$$P_{o \text{ (max.)}} = \frac{E_a^2}{2} \times \frac{2r_a}{16r_a^2} = \frac{E_a^2}{16r_a}$$

The grid bias for this output will be

$$\frac{E_a}{\mu} \times \frac{3r_a}{4r_a} = \frac{3}{4} \frac{E_a}{\mu}$$

and the efficiency

$$\frac{R_c}{2(2r_a + R_c)} = \frac{2r_a}{2(2r_a + 2r_a)} = \frac{1}{4} \text{ or 25 per cent.}$$

Advantage of high efficiency

19. Summarising the above, it may be stated that in a triode oscillator in which the anode current executes sinusoidal variations and grid current is allowed to flow during a large portion of every cycle, the efficiency may approach 50 per cent., while if grid current is totally avoided the efficiency cannot exceed 25 per cent. Under the latter conditions the maximum power output is only one-third of that expended in heating the valve, which is consequently much larger and more expensive than would be required if a higher efficiency were achieved. For example, in a low-power ground station, an output of 200 watts may be required, and if the transmitter operates at an efficiency of only 25 per cent. the input power must be 800 watts, of which 600 watts are dissipated in the valve itself. If, however, an efficiency of 80 per cent. can be achieved the input power is only 250 watts of which 200 watts are utilised and 50 watts dissipated by the valve. The economy resulting from the use of smaller and less expensive components is of far greater importance than the actual saving in electrical energy, particularly in sets designed for use in aircraft. When operating under sinusoidal conditions, the energy is supplied to the oscillatory circuit continuously, at the exact rate at which it is converted into oscillatory power. This mode of supply is not essential, for an oscillation will be maintained at a constant amplitude if the necessary amount of energy is supplied in bulk during a short interval during each period. The action is analogous to that of a child's swing which is maintained in oscillation by giving correctly timed pushes at a certain point during each oscillation, and when an electric circuit is supplied with energy in this way it is said to undergo "impulse excitation."

20. (i) High efficiency of power conversion is achieved by maintaining the grid potential at a value so negative that anode current flows during only a portion of each cycle instead of during the whole cycle as in sinusoidal operation, for less energy is then dissipated as heat at the anode. The variation of anode current during a single cycle of grid voltage, under three different conditions of grid bias, is shown in fig. 14, (a), (b) and (c). At (a) the mean negative grid potential is such

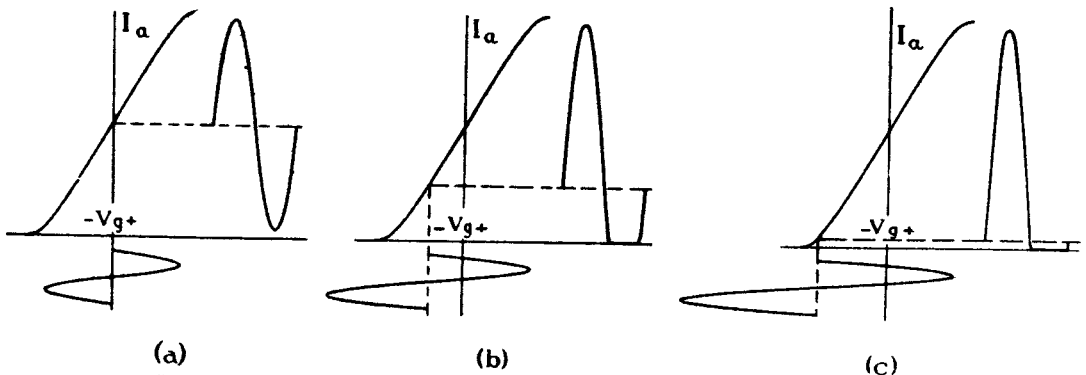


FIG. 14, CHAP. IX.—Reduction of mean anode current by negative grid bias.

CHAPTER IX.—PARA. 21

that the mean anode current is one-half the saturation value, while the amplitude of oscillatory grid voltage is just sufficient to cause a variation of anode current between saturation value and zero. This is the condition under which, on the assumption that the anode-filament P.D. may be allowed to fall to zero at the instant when the oscillatory P.D. across the tuned circuit is a maximum, a theoretical efficiency of 50 per cent is attainable. At (b) the grid has been given a negative bias but no corresponding increase in anode voltage, and although the variation of grid filament potential is still sinusoidal, the corresponding changes of anode current do not obey the simple sine law. The average anode current is not one-half the saturation value, but rather less than this, and the input power is correspondingly reduced. The total change of anode current, however, is the same as in the previous instance, i.e. from its mean value to saturation, then falling to zero and finally returning to its mean value, in the course of one cycle of oscillatory grid voltage, and consequently the power transferred to the oscillatory circuit is the same in each case, and the latter mode of operation results in the same output with a reduced input power, that is, higher efficiency of power conversion.

(ii) At (c) the grid has been given such a large negative bias that an appreciable anode current flows only during the positive half-cycle of grid voltage, and the corresponding change of anode current takes the form of a half sine wave, approximately. Under this operating condition the input power is still further reduced, while the power converted into oscillation is practically the same as before, hence the efficiency is higher than under the conditions shown in fig. 14b. It will be shown that under the operating conditions of fig. 14c a theoretical efficiency of 78.5 per cent. is obtained. It must always be borne in mind that as the negative bias is increased it is necessary to increase the peak value of oscillatory grid voltage in order that the maximum possible variation of anode current may be obtained, otherwise the power output will fall off as the efficiency is increased. The total excursion of grid voltage is obviously equal to twice its peak value, and this is generally referred to as the "grid swing." It will be seen in fig. 14 that the operating condition at (c) requires twice the grid swing called for by the mode of operation shown at (a), but this increase in oscillatory grid voltage is easily attained by increasing the coupling between the anode and grid circuits. As the peak anode current must at least approach saturation value, it is essential that the grid voltage shall be allowed to assume positive values during a portion of the cycle, otherwise the output will be low. It would seem that this necessity could be avoided by increasing the anode supply voltage and so ensuring that saturation current occurs at the instant at which the grid voltage is zero, but with most transmitting valves this would entail the employment of an excessively high anode voltage. In general it may be said that maximum output can only be obtained by allowing the valve to pass grid current during some portion of the oscillatory cycle.

Efficiency of oscillator biased to "cut off" point

21. The efficiency and output power under conditions corresponding to fig. 14c are approximately determined as follows:—It is assumed that the $I_a - V_a$ curves have the ideal form shown in fig. 15, and in the preliminary stage we may suppose it is permissible to allow the anode filament P.D. to fall to zero at the instant of peak oscillatory voltage across the output circuit. Let A B be the load line for an effective external load R_e and O B the steady anode supply voltage E_a which is equal to the peak value V_a of the oscillatory voltage across the load. The peak value of anode current is O A = I_p . The negative grid bias is just sufficient to ensure that anode current flows only on positive half-cycles of grid-filament voltage, and the anode current therefore takes the form of impulses of semi-sinusoidal form as previously explained. The analysis of such a wave-form shows that it consists of a fundamental component $\mathcal{I}_a$ the amplitude of which is one half the peak value, or $\frac{I_p}{2}$, together with a series of even harmonics. Only the fundamental will set up an appreciable oscillatory voltage across the anode circuit because it is tuned to the fundamental frequency, and this voltage wave-form is practically sinusoidal in spite of the presence of harmonics in the current wave. The mean value of the anode current over a complete cycle is $\frac{I_p}{\pi}$, and the input power is $\frac{E_a I_p}{\pi}$. The output power will be equal to the product of the

R.M.S. current and voltage across the tuned circuit. The R.M.S. current is $\frac{I_p}{\sqrt{2}} = \frac{I_p}{2\sqrt{2}}$ and the R.M.S. voltage is $\frac{V_a}{\sqrt{2}}$; the output is thus equal to $\frac{E_a I_p}{4}$ and the efficiency is

$$\eta = \frac{\frac{E_a I_p}{4}}{\frac{E_a I_p}{\pi}} \times 100 = \frac{\pi}{4} \times 100 = 78.54 \text{ per cent.}$$

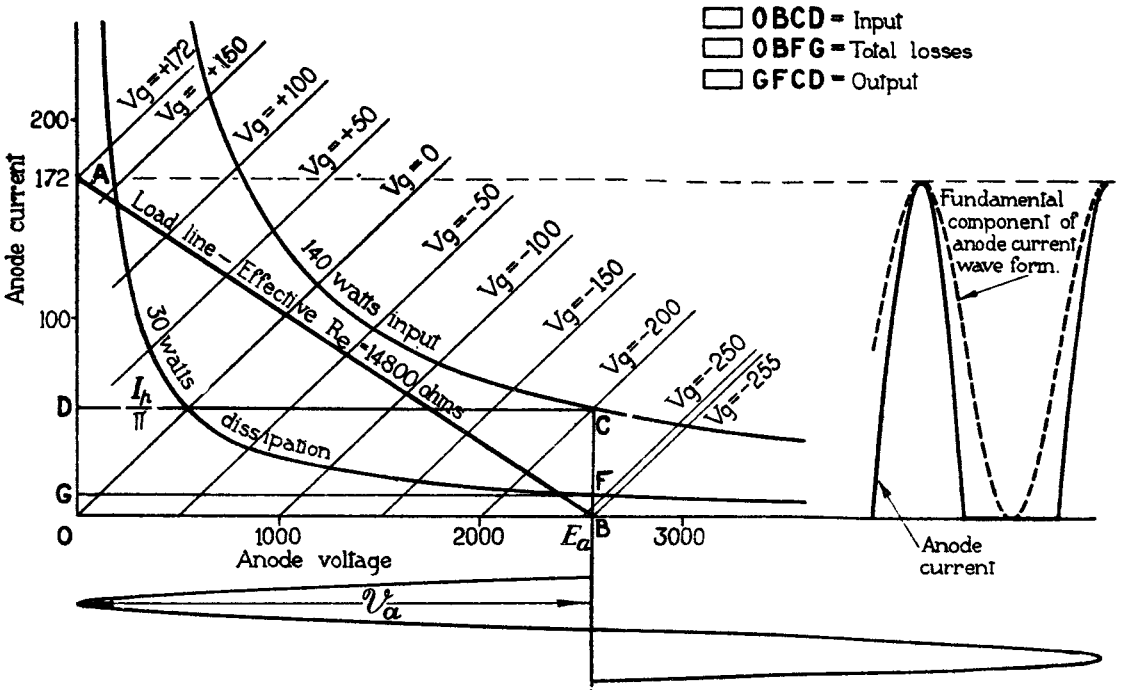


FIG. 15, CHAP. IX.—Theoretical efficiency and output with impulse excitation.

If the output is to be the greatest possible without exceeding the permissible power dissipation of the valve, the effective dynamic resistance of the load is also fixed. The actual dynamic resistance R_d is obviously equal to $\frac{V_a}{I_p} = \frac{2E_a}{I_p}$ but its effective value R_e is only $\frac{E_a}{I_p}$ because so far as the valve is concerned current only flows during alternate half-cycles. The anode dissipation is equal to the difference between input and output power and must be caused by the mean steady current and the D.C. resistance of the valve. However, with the ideal characteristics postulated, this is identical with the anode A.C. resistance. Algebraically, then,

$$\frac{E_a I_p}{\pi} - \frac{E_a I_p}{4} = \left(\frac{I_p}{\pi} \right)^2 r_a.$$

CHAPTER IX.—PARA. 21

Dividing throughout by I_p^2

$$\frac{E_a}{\pi I_p} - \frac{E_a}{4I_p} = \frac{r_a}{\pi^2}.$$

As $\frac{E_a}{I_p} = R_e,$

$$\begin{aligned} R_e \left(\frac{1}{\pi} - \frac{1}{4} \right) &= \frac{r_a}{\pi^2} \\ R_e &= \frac{4}{\pi} \frac{r_a}{4 - \pi} \\ &= 1.48 r_a. \end{aligned}$$

It must be borne in mind that the dynamic resistance of the load is twice this, or $2.96 r_a$. In fig. 15 the "virtual load line" A B is drawn with a slope corresponding to $1.48 r_a$.

The peak value of the anode current is dependent solely upon the permissible anode dissipation P_L and the resistance of the valve.

$$\begin{aligned} P_L &= \left(\frac{I_p}{\pi} \right)^2 r_a \\ I_p &= \pi \sqrt{\frac{P_L}{r_a}} \end{aligned}$$

while the peak oscillatory voltage $\mathcal{V}_a$ and the supply voltage E_a are obtainable by the relations

$$\mathcal{V}_a = \mathcal{I}_a R_d$$

$$E_a = I_p R_e.$$

Example 1.—The $I_a - V_a$ curves of fig. 15, are those of the V.T.1A valve; $P_L = 30$ watts, $r_a = 10,000$ ohms, $\mu = 10$. The optimum (virtual) load is 1.48×10^4 ohms and the peak anode current $= \pi \sqrt{\frac{30}{10^4}}$ amperes or 172 milliamperes.

$$E_a = .172 \times 1.48 \times 10^4 = 2,550 \text{ volts.}$$

$$\text{Input} = \frac{E_a I_p}{\pi} = 140 \text{ watts.}$$

$$\text{Output} = \frac{E_a I_p}{4} = 110 \text{ watts.}$$

$$\text{Anode dissipation} = 140 - 110 = 30 \text{ watts.}$$

These are the conditions shown in fig. 15. The input grid-filament voltage may be calculated as follows. The peak anode current is that which would be obtained in an equivalent circuit consisting of a generator of $\mu \mathcal{V}_g$ volts having an internal resistance r_a and a load resistance R_e . Hence

$$\begin{aligned} I_p &= \frac{\mu \mathcal{V}_g}{r_a + R_e} \\ &= \frac{\mu \mathcal{V}_g}{2.48 r_a} \end{aligned}$$

As

$$I_p = \pi \sqrt{\frac{P_L}{r_a}}$$

$$\pi \sqrt{\frac{P_L}{r_a}} = \frac{\mu \mathcal{V}_g}{2.48 r_a}$$

or

$$\mathcal{V}_g = \frac{2.48 \pi}{\mu} \sqrt{P_L r_a}.$$

In the chosen example,

$$\mathcal{V}_g = \frac{2 \cdot 48\pi}{10} \sqrt{30 \times 10^4}$$

$$= 426 \text{ volts.}$$

In fig. 15 the grid bias is -255 volts, while at the anode current peak the grid voltage reaches $+172$ volts, and the amplitude of the grid-filament oscillatory voltage must be $172 + 255 = 427$ volts. The slight divergence between graphical and calculated values is of no practical significance.

22. It is necessary in practice to stipulate that the anode-filament P.D. shall not fall below a certain value, say E_o volts. The supply voltage must then be $E_a = E_o + \mathcal{V}_a$ and the power input $\frac{(E_o + \mathcal{V}_a) I_p}{\pi}$ watts. The total losses will be increased because an amount of power $\frac{E_o I_p}{\pi}$ will be dissipated in addition to the quantity $\left(\frac{I_p}{\pi}\right)^2 r_a$.

Hence
$$\frac{r_a}{\pi^2} I_p^2 + \frac{E_o}{\pi} I_p = P_L$$

or
$$\frac{r_a}{\pi^2} I_p^2 + \frac{E_o}{\pi} I_p - P_L = 0.$$

Solving this as a quadratic equation we find

$$I_p = -\frac{E_o \pi}{2r_a} + \sqrt{\left(\frac{E_o \pi}{2r_a}\right)^2 + \frac{P_L \pi^2}{r_a}}.$$

The positive sign only is inserted before the radical because I_p must be a positive quantity. The power balance sheet now becomes

$$\text{Input} = (\mathcal{V}_a + E_o) \frac{I_p}{\pi}$$

$$\text{Output} = \frac{\mathcal{V}_a I_p}{4}$$

$$\text{Losses} = \left(\frac{I_p}{\pi}\right)^2 r_a + \frac{E_o I_p}{\pi}$$

$$\text{Input} - \text{Losses} = \frac{\mathcal{V}_a I_p}{\pi} + \frac{E_o I_p}{\pi} - \frac{I_p^2 r_a}{\pi^2} - \frac{E_o I_p}{\pi}$$

$$\text{Output} = \frac{\mathcal{V}_a I_p}{\pi} - \frac{I_p^2 r_a}{\pi^2}$$

$$\text{Hence} \quad \frac{\mathcal{V}_a I_p}{\pi} - \frac{\mathcal{V}_a I_p}{4} = \frac{I_p^2 r_a}{\pi^2}$$

which signifies that the optimum virtual load resistance is independent of E_o , and is in fact $1 \cdot 48 r_a$ as before. Thus if it is necessary to limit the minimum anode-filament P.D. to about 250 volts (say 80π or $251 \cdot 3$ volts)

$$I_p = -\frac{80\pi^2}{2 \times 10^4} + \sqrt{\left(\frac{80\pi^2}{2 \times 10^4}\right)^2 + \frac{30\pi^2}{10^4}}$$

$$= 134 \text{ milliamperes.}$$

$$\mathcal{V}_a = I_p R_e = 134 \times 14,800 = 1,985 \text{ volts.}$$

$$E_a = \mathcal{V}_a + E_o = 1,985 + 251 \cdot 3 = 2,236 \text{ volts.}$$

CHAPTER IX.—PARA. 22

$$\text{Input power} = \frac{E_a I_p}{\pi} = \frac{2,236 \times .134}{\pi} = 95.4 \text{ watts}$$

$$\text{Output power} = \frac{r_a I_p}{4} = \frac{1,985 \times .134}{4} = 66.5 \text{ watts}$$

$$\text{Total losses} = 95.4 - 66.5 = 28.9 \text{ watts}$$

$$\text{Efficiency} = \frac{\text{output}}{\text{input}} \times 100 = \frac{6,650}{95.4} = 69.7 \text{ per cent.}$$

The required grid-filament input voltage is found as before and is 332 volts, while the grid bias is $-\frac{E_a}{\mu}$ or -223.6 volts. Hence the grid will now swing positive up to 110 volts as shown in fig. 16.

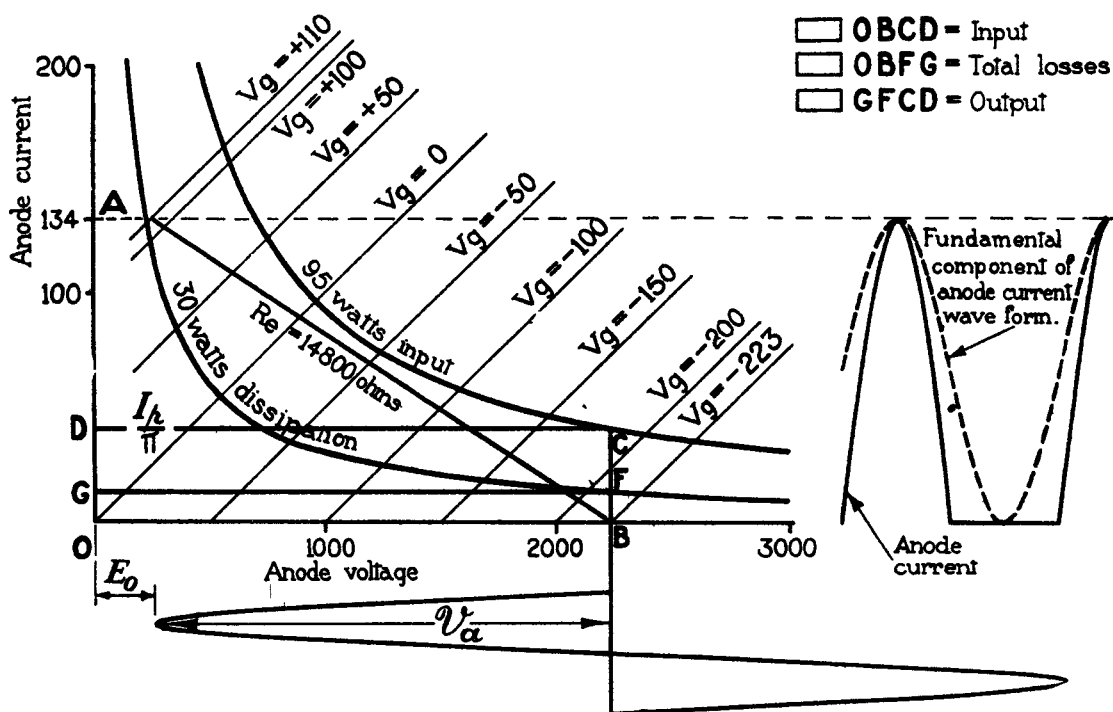


FIG. 16, CHAP. IX.—Efficiency and output with impulse excitation.

It may here be observed that in many instances the emission available from the filament may be less than that required to give both maximum efficiency and output. If the emission of the valve in the previous examples is only 100 milliamperes, and the available H.T. voltage 1,000 volts, the maximum input power under semi-sinusoidal conditions will be 31.8 watts, and at 70 per cent. efficiency the output will be about 22 watts, hence the power dissipated in the valve will only be about 10 watts instead of 30. Again, if the valve has ample emission but the H.T. voltage is limited owing to questions of safety or insulation of various components (the latter factor being of importance when space is limited as in aircraft transmitters) the optimum load impedance for maximum output will often be less than the anode A.C. resistance of the valve.

Production of harmonics

23. The increased efficiency of power conversion obtainable by impulse excitation is accompanied by the production of harmonic variation of anode current, that is to say, the non-sinusoidal variation of anode current is really the sum of a number of sinusoidal variations and a mean steady current. These harmonic variations of anode current produce oscillatory currents of corresponding frequency in the oscillatory circuit, but the proportion of harmonic to fundamental oscillatory current is less than the corresponding proportion in the anode current, because the anode circuit is tuned to the fundamental frequency but not to the frequency of the harmonic. When direct aerial excitation is employed, the proportion of second harmonic to fundamental aerial currents is given by the following formula.

Let I_1 be the component of anode current at fundamental frequency,

I_2 the component of anode current at second harmonic frequency,

χ the magnification of the aerial circuit,

$\mathcal{I}_1$ the component of aerial current at fundamental frequency,

$\mathcal{I}_2$ the component of aerial current at second harmonic frequency,

then
$$\frac{\mathcal{I}_2}{\mathcal{I}_1} = \frac{I_2}{I_1} \times \frac{4}{3\chi}$$

Although impulse excitation gives higher efficiency than sinusoidal operation, the power converted into oscillations at harmonic frequencies is wasted so far as the distant receiver is concerned, although this radiation can and does cause interference with receivers in the neighbourhood of the transmitter. In order to avoid this interference, some form of indirect coupling between aerial and oscillatory circuits may be adopted. In such instances it is preferable to couple the aerial to the inductive branch of the oscillatory circuit rather than to the capacitive branch, for the harmonic currents are always greater in the latter path than in the former.

Methods of obtaining grid bias in transmitter

24. The grid bias may be provided by a battery or by a motor generator, the latter method being rarely adopted. Battery bias is sometimes used in low power C.W. transmitters and in the amplifier stages of frequency-controlled transmitters. Frequently, however, the bias is obtained by allowing grid current to flow during a portion of the cycle. In series with the grid reaction coil is placed a condenser, in parallel with which is a resistance of the order of 10,000 ohms. The action of this condenser may be best studied by first considering the resistance to be absent, and assuming oscillation to commence. On completing the filament circuit, but with the anode circuit open, the grid may possibly collect a few electrons if it possesses some slight positive potential with respect to the filament, but even so, in a very short time its potential must become equal to that of the filament and it will be assumed that this is the case when the anode circuit is completed. The production of oscillations is accompanied by variation of grid potential as previously described, and whenever the grid is positive with respect to the filament, it acts as a collecting as well as a controlling electrode. The electrons collected by the grid will charge the grid condenser, negatively on the plate connected to the grid, and a corresponding positive charge will be developed on the plates connected to the filament. The grid thus acquires a potential negative to the filament which becomes progressively greater in each succeeding cycle and the anode current is correspondingly reduced. With sufficient coupling between grid and anode circuits, the grid becomes so negative that the anode current is reduced to zero. The oscillation then dies away and does not recommence, because the grid condenser retains its charge and prevents the re-establishment of anode current.

25. Now consider the action with a "grid leak" of large resistance connected in parallel with the condenser. When oscillations commence, the grid collects electrons on every positive half-cycle of grid voltage, but these electrons can now leak away to an extent depending upon the resistance of the leak. With a suitable value of the latter the grid potential gradually becomes more negative, just as before, but eventually reaches some value at which the number of electrons collected by the grid during the portion of the cycle in which grid current is flowing is just equal

CHAPTER IX.—PARAS. 26-27

to the number which escape from the grid through the resistance during the whole cycle. The grid potential then remains practically constant at this value, which can be adjusted to any desired amount by choice of capacitance of grid condenser, resistance of leak, and coupling between grid and anode circuits. It is found that satisfactory operation can be obtained in nearly all transmitters if the capacitance of the grid condenser is from $\cdot 0001$ to $\cdot 002 \mu F$ (the higher values being used for low frequency transmitters) and the leak resistance from 10,000 to 50,000 ohms. The condenser and leak method of obtaining grid bias possesses an important advantage over all other methods, for with this device oscillations always commence with zero grid bias, and the initial anode current variations occur on the steepest portion of the valve characteristic, resulting in a rapid growth in the amplitude of the oscillation. As already stated, by correct choice of value of grid condenser and leak resistance, the grid bias will increase in value as the amplitude of oscillation increases. If some permanent form of grid bias is employed, such as a battery, the mean grid voltage is the same over the period during which the amplitude of oscillation is growing as when the final amplitude is reached, and consequently the initial changes of anode current occur near the curved foot of the characteristic where the slope is small. Under such conditions considerable time is taken for the oscillation to reach maximum amplitude, and too great a value of negative bias may prevent the inception of oscillations.

Intermittent oscillations

26. When the grid bias is derived by the condenser and leak method, it is sometimes found that the oscillations are periodically interrupted, either at an audible rate or possibly several thousand times per second. Intermittent oscillation occurs when the coupling between the anode and grid circuits is much greater than is necessary for the bare maintenance of oscillations, and the time constant $C_g R_g$ of the condenser and leak combination is so large that the charge on the grid condenser cannot leak away with sufficient rapidity. Oscillations start as usual about an operating point situated upon the linear portion of the I_a-V_g characteristic, but the operating point moves downward and to the left into the region where the dynamic mutual conductance is low, and the amplitude of the oscillation starts to decrease. With the correct value of grid leak and degree of coupling, the decrease of amplitude is followed almost instantaneously by a change of bias, stabilising the amplitude of the oscillation at this value as already stated. Under the conditions now in question, however, the decrease of amplitude is not immediately followed by a decrease of negative bias owing to the slow rate of discharge of the grid condenser, and consequently the amplitude of the next oscillation is still further reduced, the effect being cumulative. When the oscillation has died out, the grid condenser still continues to discharge, slowly, through the leak, and the oscillations recommence when the operating point has reached a location on the I_a-V_g characteristic at which the conditions are again favourable. This phenomenon is occasionally utilised in the design of low power oscillators from which a Type A2 (Tonic Train) emission is desired; such oscillators are often referred to as "squeggers." Its occurrence in a C.W. transmitter indicates a breakdown of the grid condenser or grid leak resistance, calling for immediate attention.

Function of the anode tapping point

27. As the output circuit is in effect a rejector circuit in series with the valve, the latter being considered to act as an A.C. generator of voltage μV_g at the resonant frequency of the output circuit, it seems at first sight that the scheme of connections shown in fig. 1 would be perfectly satisfactory for all purposes. Unfortunately, however, the maximum oscillatory power is rarely developed under these circumstances for the following reasons. The value of the product LC in the oscillatory circuit is fixed by the frequency which the transmitter is to radiate, because

$$f = \frac{1}{2\pi\sqrt{LC}}.$$

Taking L and C independently, the capacitance is usually that of the aerial alone,

its value depending upon the design of the aerial and therefore upon the circumstances in which the transmitter is to be used, i.e. air or ground. The capacitance of the aerial being practically incapable of variation, the adjustment of frequency is generally made by varying the inductance in series with the aerial. On the other hand if maximum output power is required, we have seen

that the dynamic resistance of the oscillatory circuit must bear a definite relation to the internal resistance of the valve, e.g. for sinusoidal operation without grid current, $R_d = 2r_a$. Now as $R_d = \frac{L}{CR}$ and the resistance of the aerial circuit is also constant at any given frequency, it appears that the correct loading conditions can only be obtained by a suitable choice of the ratio $\frac{L}{C}$, but as both L itself and the product LC are already fixed by other considerations, it is usually impossible to vary this ratio. The desired effect can be achieved however, by connecting the valve to the output circuit in the manner shown in fig. 17a. Here the circuit consists of two parallel branches, one containing an inductance L only, the other containing an inductance l and a condenser C in series. The circuit must of course possess an inherent resistance R , which may be considered as "lumped" at any point in the circuit, its exact location being immaterial because it is so small that its effect on the phase of the current is negligible. The reactance of the purely inductive branch is ωL ohms, and that of the other branch is $\frac{1}{\omega C} - \omega l$ ohms. Hence,

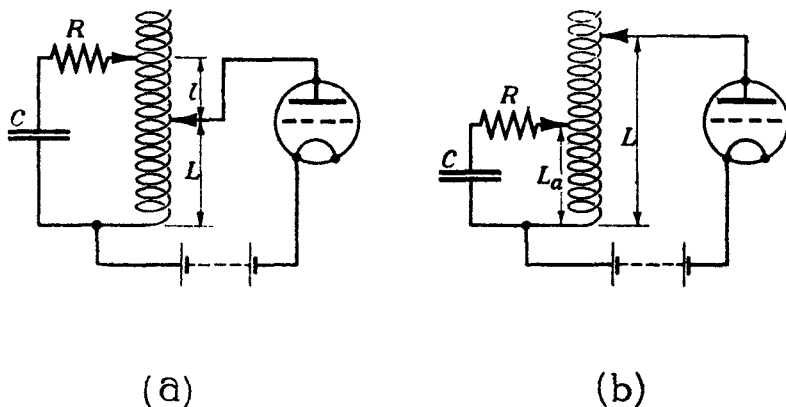


FIG. 17, CHAP. IX.—Anode tapping point.

if a voltage V is applied to the circuit

$$I_L = \frac{V}{\omega L} \text{ lagging on } V \text{ by } 90^\circ,$$

$$I_C = \frac{V}{\omega C - \frac{1}{\omega l}} \text{ leading on } V \text{ by } 90^\circ,$$

the effect of the resistance being neglected.

The supply current I_a is equal to the arithmetical difference of these or

$$I_a = V \left(\frac{1}{\omega L} - \frac{1}{\omega C - \frac{1}{\omega l}} \right)$$

The supply current will be zero if $\frac{1}{\omega L} = \frac{1}{\omega C - \frac{1}{\omega l}}$ or $\omega L = \frac{1}{\omega C} - \omega l$, that is if $\omega^2 = \frac{1}{(L + l)C}$.

It will be noted therefore that in a rejector circuit of total inductance $L + l$ and capacitance C the resonant frequency is not changed by altering the position of the two points to which the supply voltage is connected. The circulating current I_L or I_C for a given supply voltage is however

CHAPTER IX.—PARA. 28

greater in the circuit of fig. 17a than in the original rejector circuit, for in the latter the current in the inductive branch for example, would be $\frac{V}{\omega(L+l)}$ instead of $\frac{V}{\omega L}$ as in the present condition.

28. The power P expended in the oscillatory circuit depends upon the resistance, and is equal to $I_L^2 R$ watts. Since $I_L = \frac{V}{\omega L}$,

$$P = \frac{V^2}{(\omega L)^2} R = \frac{V^2 C R (L+l)}{L^2} \text{ watts}$$

because $\omega^2 = \frac{1}{(L+l)C}$. An alternative expression for the power expended is $\frac{V^2}{R_d}$ where R_d is the dynamic resistance of the parallel circuit, and equating these

$$\begin{aligned} \frac{V^2}{R_d} &= \frac{V^2 C R (L+l)}{L^2} \\ R_d &= \frac{L^2}{(L+l)CR} \\ &= \left(\frac{L}{L+l}\right)^2 \frac{L+l}{CR} \text{ ohms.} \end{aligned}$$

Now $\frac{L+l}{CR}$ is the dynamic resistance of the tuned circuit when connected in the anode circuit with no anode tap, and when the latter device is fitted, the dynamic resistance may be either larger or smaller than this. The physical explanation is as follows. The aerial tuning inductance may be considered to act as an air-core auto-transformer, and in Chapter VI it is shown that if the resistance of the secondary circuit of a transformer is R_s ohms, the equivalent resistance transferred to the primary circuit is $\frac{R_s}{T^2}$ ohms, T being the transformation ratio. The effective

transformation ratio in the present instance is $\frac{L+l}{L}$, and as the value of L may be varied by means of the anode tap without affecting the value of $L+l$, the effective resistance transferred to the anode circuit may also be varied. This effect is easily illustrated by numerical examples.

Example 2.—A valve of $r_a = 10,000$ ohms has an anode circuit consisting of an inductance of $2,000 \mu H$, and a capacitance of $0.001 \mu F$ in parallel, the total resistance of the oscillatory circuit being 50 ohms. The conditions of operation require that R_d shall be $2r_a$. What portion of the total inductance should be included in the anode circuit?

The dynamic resistance of the anode circuit without anode tap is $\frac{2,000}{0.001 \times 50}$ or 40,000 ohms, and for maximum output this must be reduced to 20,000 ohms.

$$\begin{aligned} \frac{40,000}{20,000} &= \left(\frac{L+l}{L}\right)^2 \\ \frac{L+l}{L} &= \sqrt{2} \\ L &= 0.707 (L+l) \end{aligned}$$

i.e. the anode tapping point should be adjusted so that 0.707 of the aerial tuning inductance is included in the anode circuit. At high frequencies, it may be necessary to use a step-down auto-transformer in order to secure correct loading conditions, as in the following example.

Example 3.—If the valve and circuit remain as above, except that the value of aerial inductance required for tuning purposes is $200 \mu H$, what is the position of the anode tap for maximum output?

Instead of denoting the tuning inductance by $L+l$ it will be preferable to denote it by a single symbol, L_a .

The dynamic resistance of the oscillatory circuit is now $\frac{L_a}{CR} = \frac{200}{.001 \times 50} = 4,000$ ohms.

$$\left(\frac{L_a}{L}\right)^2 = \frac{4,000}{20,000}$$

$$L = \sqrt{5} L_a \\ = 2.24 L_a.$$

This signifies that 2.24 times as much inductance must be included in the anode circuit as is required in the oscillatory circuit, and the arrangement used is shown in fig. 17b. This condition only arises in practice in the high frequency band.

29. When the anode tap has been adjusted for maximum output the external and internal impedances R_d and r_a are said to be matched, and it must be clearly understood that by matching, equality of internal and external impedance is not necessarily implied. In practical transmitters two methods of adjusting the anode tap may be met with. In large transmitters designed for use on the ground, the aerial capacitance at one station may differ considerably from that at another, and as space and weight are not of primary importance the design of the transmitting inductance may be sufficiently flexible to allow the desired frequency range to be covered with any aerial system likely to be adopted. In such circumstances the anode tap is independently adjustable, its optimum position being found by trial and error. In aircraft, however, the aerial design for use with a given type of transmitter is standardised, and it is generally possible to design the transmitter in such a manner that the operation of changing the value of tuning inductance automatically adjusts the anode tap to its optimum value. The advantage of this is the saving of time which might otherwise be spent in an endeavour to achieve a slight improvement in output power.

The pulse coil

30. In certain transmitters operating upon the high frequency band, the tuning inductance consists only of a variometer. Owing to the necessity for relative movement of its coils it is not desirable to arrange an anode tapping point directly upon the variometer windings. The matching of internal and external impedances is then achieved by the addition of what is called a pulse coil, the arrangement being shown in fig. 18a and 18b. It will be seen that the pulse coil is in effect an auto-transformer which may be used either as a step-up, step-down or unity transformation ratio device by judicious selection of the tapping points, and matching can be achieved over a wide frequency range. The inductance of the pulse coil should be of the order of ten times the maximum inductance of the variometer winding.

31. In designing a transmitter, the power output required may be taken as the starting point, and together with the efficiency of power conversion, a figure for which may be assumed

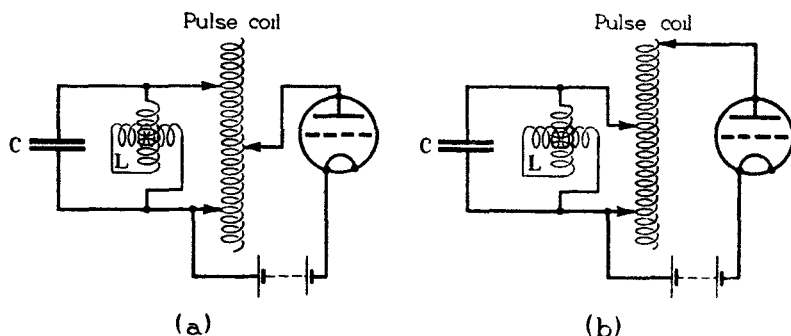


FIG. 18, CHAP. IX.—Use of pulse coil.

CHAPTER IX.—PARAS. 32-33

from experience of similar designs, the rating of the valve is obtained ; for example if the required output is 100 watts and an efficiency of 66 per cent. is considered feasible, 100 watts will be two-thirds of the input power, hence the latter must be 150 watts, and the valve is required to dissipate one-third of the input, or 50 watts. The ratings of both the H.T. generator and the valve are now known, and the remainder of the circuit is built up round these fundamental components. As it is impossible to provide an indefinite range of valves of various ratings, it is often necessary to use two or more valves in order to obtain the desired output, the valves being connected in series, in parallel, or in push-pull. The first method offers practical difficulties which prevent its adoption, but the two latter methods are in general use.

Valves in parallel

32. Valves are said to be in parallel when the corresponding electrodes are interconnected, the filaments being connected in parallel and fed from a common L.T. source, the anodes connected to a common terminal and thence to the anode tap, while the grids also are connected to a common terminal and thence to the reaction coil or tuned grid circuit as the case may be. If, in case of emergency, an attempt is made to increase the output of an existing transmitter by connecting valves in parallel, it is important to observe that the valves must be as nearly identical as possible, otherwise there is a possibility that one or more may act merely as shunt resistances across the others, absorbing power instead of supplying it to the oscillatory circuit. This is not likely to happen unless valves of very different characteristics, e.g. a V.T. 25 and a V.T. 13C are so connected, but the permissible variation decreases as the number of valves is increased. As the anode A.C. resistance of two similar valves in parallel is only one-half the resistance of a single valve, the optimum external resistance will be lower, and the amount of inductance included between the anode tap and the low potential end of the anode inductance must also decrease. This requirement places a limit on the number of valves which may be operated in parallel, for a point is reached at which the inductance included in the anode circuit becomes so small that the phasing conditions necessary to maintain oscillation are no longer preserved. If two valves are perfectly matched, and are working into the correct load resistance, the power output should be double that obtainable from a single valve under the same operating conditions. The circulating current in the oscillatory circuit will therefore be increased in the ratio $1 : \sqrt{2}$.

Valves in push-pull

33. The scheme of connections given in fig. 19 is called push-pull connection. The name is perhaps unfortunate because it tends to cause the reader to expect an explanation of its operation in terms which interpret the idea of pushing and pulling too literally. Generally, valves may

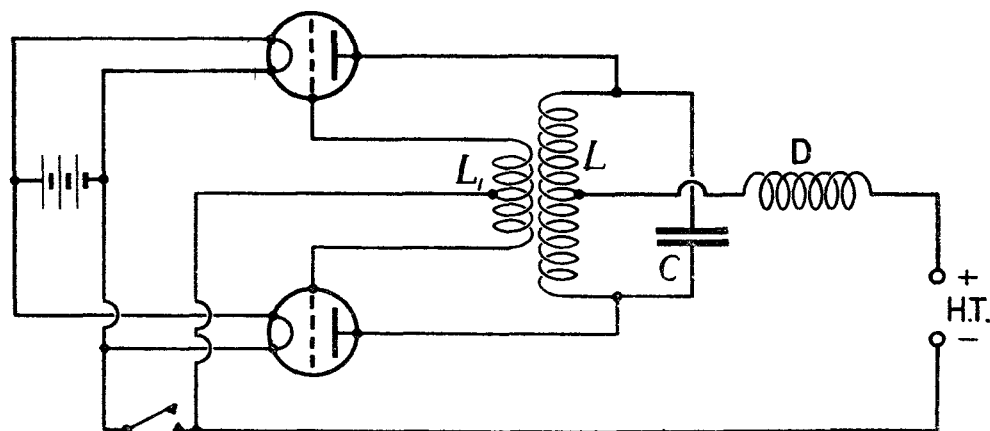


FIG. 19, CHAP. IX.—Simple push-pull oscillator.

be said to be in push-pull when one half of the total grid-filament voltage is applied to each valve of a pair, the output voltages being combined in such a manner that the total is double that of a single valve. The manner in which an oscillation may be maintained, when once established in the output circuit, is as follows :—

(i) Consider the cycle to commence at the instant at which the whole of the oscillatory energy is stored in the anode circuit inductance, and the circulating current is a maximum. At this moment the electron flow through the inductance is in the direction B to A as shown by the arrow in fig. 20a, and the induced E.M.F. in the reaction coil L_1 is zero because the connections are so arranged that the relative phases of current and voltage are as shown in fig. 3, i.e., the grid-filament voltage is 90° ahead of the current in the anode circuit inductance. The anode currents in both valves are of equal magnitude as indicated by the length of the arrows. It should be noted that the arrows indicate the direction of the electron flow in all cases. During the time the current is flowing in the inductance a certain amount of energy is dissipated in the

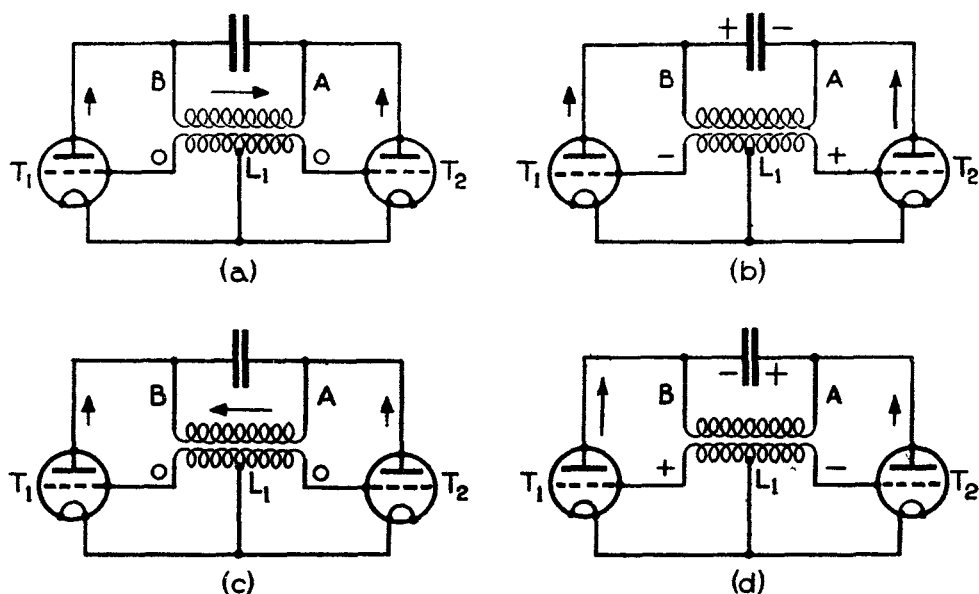


FIG. 20, CHAP. IX.—Action of push-pull oscillator.

form of heat and electro-magnetic radiation and it is convenient to consider that a certain number of the electrons which collectively comprise the oscillatory current are brought to rest in giving up this energy.

(ii) A quarter of a cycle later the energy is all stored in the condenser which is charged with the polarity shown in fig. 20b. The current in the inductance has by this time fallen to zero but the induced E.M.F. in the reaction coil is at its maximum value, the grid of the valve T_2 being positive with respect to the filament and that of the valve T_1 being negative. The anode current of T_2 is therefore increasing and a supply of electrons is introduced into the right-hand side of the condenser in addition to those carried by the circulating current previously mentioned. These electrons serve to replace those brought to rest by the conversion of the kinetic energy into various forms of wastage. The valve T_1 also assists in the re-charging of the condenser, for the reduction of its anode current implies that fewer electrons are arriving at the left-hand side of the condenser than when normal anode current is flowing.

(iii) The condenser will now commence to discharge through the inductance, electrons flowing through the coil from A to B, fig. 20c. At the instant when this current reaches its

CHAPTER IX.—PARAS. 34-36

maximum the grid-filament voltage of both valves is zero and anode currents normal. The whole of the energy is then stored in the inductance, the condenser voltage being zero. The current now continues to flow, charging the condenser negatively on the left-hand side; as the condenser voltage rises the current falls and is zero at the instant when the condenser charge is a maximum. The grid-filament voltage of the valve T_1 reaches its maximum positive value at this instant while that of T_2 is at its maximum negative value. The anode current of T_1 therefore increases so introducing a negative charge in the left-hand side of the condenser, which replaces those electrons which have given up their kinetic energy. The condenser then discharges once more, the conditions at the end of the next quarter of a cycle being those of fig. 20a, and the foregoing cycle is then repeated.

34. The manner in which the inception of the oscillation takes place is not obvious at first sight. In fig. 19, if the positive anode supply is connected to the electrical centre of the inductance L the two halves of the circuit are perfectly symmetrical. With valves of exactly similar characteristics, the closing of the circuit will cause equal rates of increase of anode current through both valves simultaneously and it would appear that two equal and opposite counter-E.M.F.'s are developed in the inductance, the resulting condenser charge being zero. It must be emphasized that it is not possible to achieve this exact symmetry. Even if momentarily obtained, the condition would not persist, e.g. even a minute change in the magnetic field of the earth will not affect both halves of the circuit at the same instant and to the same degree and therefore will cause some slight counter-E.M.F. to be set up, which must introduce a charge into the condenser C . The slightest asymmetry in the circuit will cause a larger counter-E.M.F. in one half of the inductance than in the other, and a resulting condenser charge, when the circuits are first completed, and once energy is stored an oscillation will occur (provided the damping is sufficiently low); the foregoing action shows that once established it will be at least maintained, and will increase in amplitude until limited by characteristic curvature if the mutual inductance between the coils L and L_1 exceeds the critical value. The impossibility of achieving exact similarity in both halves of the circuit is shown by the practical necessity for the anode choke (D, fig. 19). If absolute symmetry were attainable, this choke could be omitted, for the electrical centre of the inductance L is at the same oscillatory potential as the filament. The choke is inserted because it is practically impossible to find the exact electrical centre, and its inclusion ensures that the actual point at which the H.T. supply is fed is not thereby brought to filament potential also.

35. The advantages claimed for the push-pull type of circuit are:—

(i) The greater frequency band which may be covered by a given inductance and condenser when these components constitute either the grid (input) or anode (output) circuit of the valve. The anode-filament capacitances of the two valves are in series across the tuning condenser C in fig. 19, and the total effective capacitance in parallel with C is only one-half of that which would exist if a single valve were employed. In the same way, the grid-filament capacitances are in series across the reaction coil and the total effective capacitance of this circuit is also less than with a single valve. This effect does not appear to be of great importance in circuits in which a tuning condenser is used, for as a rule the inter-electrode capacitance will form only a very small portion of the total.

(ii) The circuit is symmetrical, and consequently variations of capacitance with respect to earth have a smaller effect upon the frequency than in circuits such as the Hartley or Colpitts.

(iii) The harmonic variations of anode current produced by impulse excitation are considerably reduced. This appears to be the most important advantage of the push-pull circuit and will be further considered.

36. The arrangement of valves in push-pull is such that the half-cycles of grid voltage which cause the grid of one valve to become positive with respect to the filament, have the opposite effect upon the grid of the other valve. If it is desired to show the effect of grid voltage upon both valves, the characteristic curves must be combined as shown in fig. 21 if operation takes place with mid-point bias, and as shown in fig. 22 if cut-off is applied. In transmitters using

impulse excitation, the latter condition is applicable and it will be assumed that the operating point is near the curved "foot" of the curve, the grid voltage swing and resulting change of anode current being as shown. It will be observed that the anode current change is far from sinusoidal, and on the right-hand side of the figure the anode current has been analysed into its principal component frequencies, i.e. the fundamental and second harmonic. It will be observed that the fundamental components, if added, result in a first harmonic which has double the

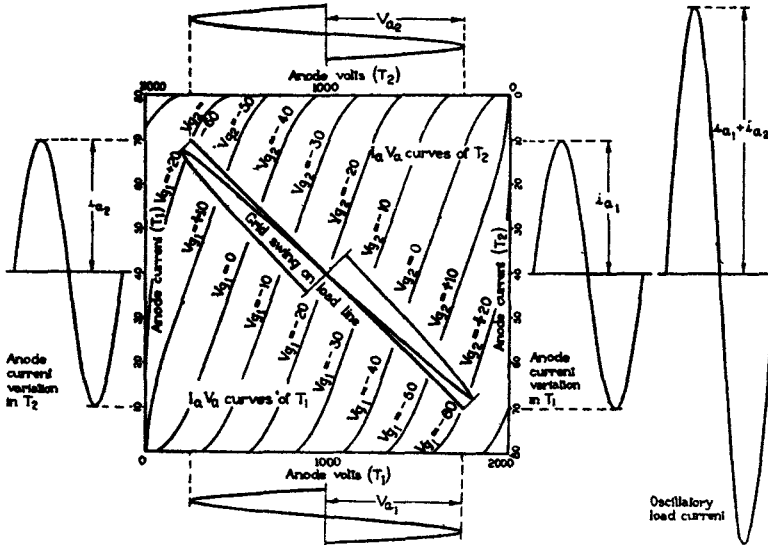


FIG. 21, CHAP. IX.—Combined characteristics of valves in push-pull with mid-point bias.

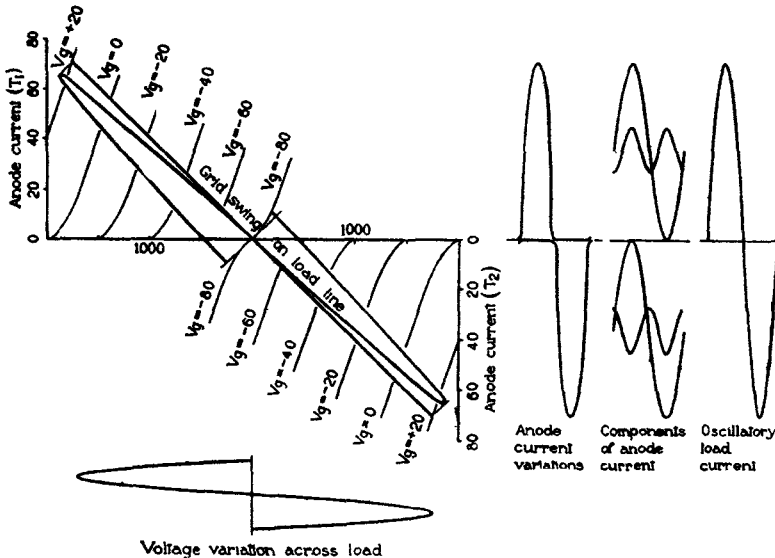


FIG. 22, CHAP. IX.—Combined characteristics of valves in push-pull, biased to foot of characteristic.

CHAPTER IX.—PARAS. 37-38

amplitude of either of the individual amplitudes, while the addition of the two second harmonic variations results in zero current. In other words, the fundamental variations of current in the oscillatory circuit are in phase with each other, while the second harmonics are in antiphase, and cancel each other. If the characteristic curves of the two valves in push-pull were identical, this cancellation of even harmonics would be complete, and therefore the system would combine the efficiency of impulse excitation with the purity of wave-form associated with operation upon the linear portion of the characteristic. Unfortunately, however, it is rarely that two valves can be so equally matched as to give more than a slight reduction of second harmonic, at any rate under service conditions. In general, it may be said that if space and weight permit the employment of two valves, it is preferable to utilise them as an oscillator and power amplifier, in the manner presently described, rather than in push-pull. The amplifier itself may however be of the push-pull type for the reasons enumerated in paragraph 35 (i) and (ii).

POWER SUPPLIES

L.T. supply to transmitter

37. In low power transmitters a secondary battery is frequently utilised to supply the necessary heating current for the valve filaments. Provided the cells are in good condition the filament current is maintained at a very steady value, and this results in an approach to stability in the values of anode A.C. resistance and grid-filament conductance of the transmitting valve or valves. The employment of battery supply for L.T. purposes is therefore conducive to a high degree of frequency constancy. Mobile transmitters e.g. in aircraft, preferably utilise battery supply for this reason, although in some instances it may be found desirable to employ a "dual purpose" generator, that is one in which two separate armatures are mounted on a common shaft, the rotor being driven through a constant speed clutch by a windmill in the slip stream. One armature then supplies say 5 amperes at 10 volts for filament heating, while the other supplies the small anode current, e.g. up to 100 milliamperes, at 2,000 volts. A machine of this kind is also a feasible method of supply to a semi-portable transmitter, the drive being supplied by a petrol motor. High power ground station transmitters may be fitted with motor generators for filament supply, owing to the reduction of maintenance compared with batteries, while at stations in which an alternating supply is available, this may be used to supply the filament circuits, through suitable step-down transformers. This generally results in a slight variation in the amplitude of the emitted radiation and a spreading of its frequency over a rather wider band than with a battery supply, but the effect is not harmful, at any rate in C.W. or I.C.W. transmission.

H.T. supply to transmitter

38. Low power transmitters may utilise a battery of inert or secondary cells, provided the total power input does not exceed a few watts, but for higher powers, either direct or alternating current generators must be employed. The direct current generator is only suitable for voltages up to about 4,000 volts, difficulties with commutation and effective insulation between armature windings and core being experienced with higher voltages. For this reason, alternating current generators are almost universally employed for high power transmitters, the necessary conversion into direct current being performed by a rectifying system. The term rectification is used to denote the conversion of an alternating into a direct current, and the commutator fitted to a D.C. generator is merely a particular form of mechanical rectifier. In the sense in which it is usually employed, however, rectification means the achievement of this conversion without aid of rotating machinery, and the term will be used with this signification in the subsequent paragraphs. A rectifier, then, may be defined as any conducting body or substance which does not obey Ohm's law, and the thermionic valve fulfils this condition as has already been shown. When the input voltage is large and a large power output is required the most suitable form of valve for rectifying purposes is the diode.

Half-wave rectifying system

39. The simplest form of rectifying circuit is that known as the half-wave rectifying system, and is shown in fig. 23. In this diagram the power supply is derived from an alternator, although in practice the ordinary commercial supply mains may be used. The supply voltage is usually from 200 to 250 volts, and is raised to say 3,000 volts by a step-up transformer. The filament of the rectifying diode may be heated by a suitable battery, but it is more convenient to utilise the A.C. supply for this purpose, and the diagram shows an L.T. transformer having a step-down of about 10 to 1, giving a filament supply at about 20 to 25 volts. An ammeter and rheostat are fitted in order that the filament current may be adjusted to the value shown on the label of the valve. So far as power supply is concerned, the transmitter may be represented by a resistance R which is connected in parallel with a condenser C called the reservoir condenser. The action of the system may be explained by first assuming that the load resistance is absent. The filament being heated to its correct temperature, an anode current will be established only when the anode is at a positive potential with respect to the filament. On closing the switch S an alternating current flows in the primary winding of the transformer and an alternating E.M.F. is

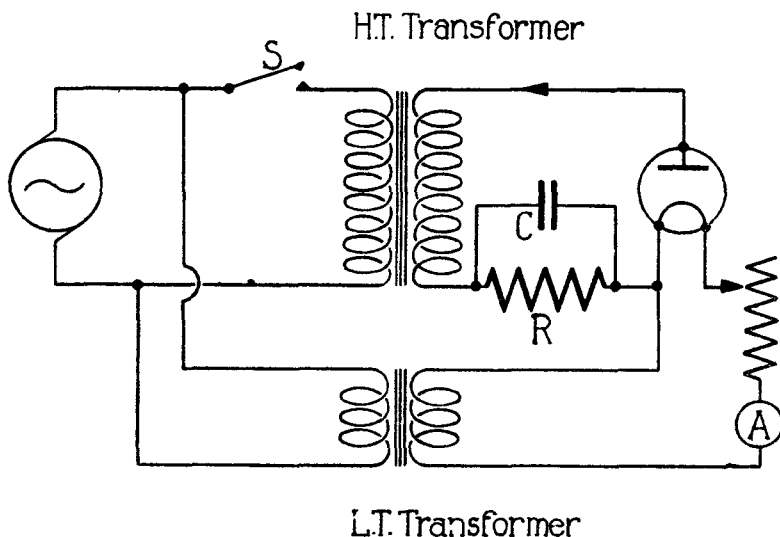


FIG. 23, CHAP. IX.—Half-wave rectifying circuit.

developed in the secondary winding. During those half-cycles in which the anode is positive an electron current will flow through the valve and the secondary winding of the transformer, charging the reservoir condenser negatively on the left-hand plate. A corresponding displacement current is set up in the dielectric and causes a repulsion of electrons from the right-hand plate which therefore acquires a positive charge. No such current will be established during those half-cycles in which the anode is negative with respect to the filament. After a few cycles, the P.D. between the plates of the condenser will be equal to the peak secondary voltage of the transformer, and no further action will take place. This charging process is shown graphically in fig. 24.

40. If the rectifier is assumed to possess an ideal characteristic, i.e. infinite resistance for negative values of applied voltage and a constant finite resistance for positive values, an idea of the charging process during the first half-cycle can be obtained by assuming that the peak current is equal to that which would be caused by the peak voltage, less that which would be caused by

CHAPTER IX.—PARA. 40

a steady voltage equal to the counter-E.M.F. of the condenser at the end of the half-cycle. If $\mathcal{I}_1$ is the peak value of the current, $\mathcal{E}$ the peak value of the applied voltage and V_1 the counter-E.M.F. of the condenser at the end of the first half-cycle,

$$\mathcal{I}_1 = \frac{\mathcal{E} - V_1}{r_a}.$$

The average value of this current will be $I_1 = \frac{\mathcal{I}_1}{K}$, K being equal to π if the current wave is semi-sinusoidal, hence

$$I_1 = \frac{\mathcal{E}}{K r_a} - \frac{V_1}{K r_a}$$

The charge Q_1 introduced into the condenser by this average current flowing for one half-cycle is $I_1 \times \frac{T}{2}$ or $\frac{I_1}{2f}$.

$$\therefore Q_1 = \frac{\mathcal{E}}{2K f r_a} - \frac{V_1}{2K f r_a}$$

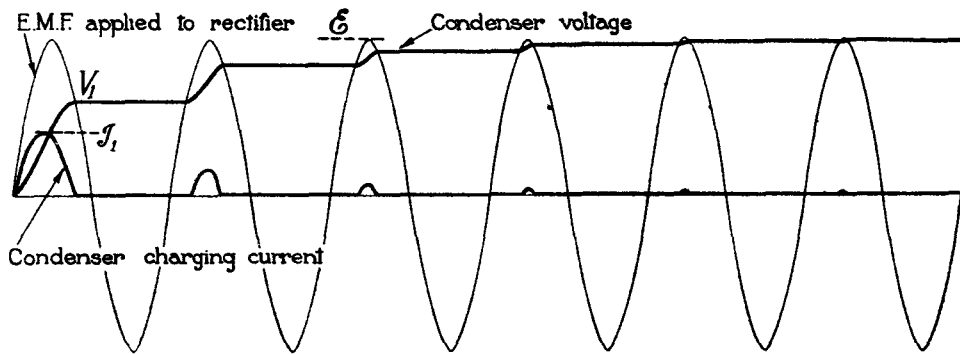


FIG. 24, CHAP. IX.—Charge of reservoir condenser under no load conditions (half-wave rectifier).

and V_1 , the voltage to which the condenser is charged, is $\frac{Q_1}{C}$, hence

$$V_1 = \frac{\mathcal{E}}{2K f r_a C} - \frac{V_1}{2K f r_a C}$$

$$V_1 \left(1 + \frac{1}{2K f r_a C} \right) = \frac{\mathcal{E}}{2K f r_a C}$$

and

$$V_1 = \frac{\mathcal{E}}{1 + 2K f r_a C}$$

As an example, take $K = \pi$, $f = 50$ cycles per second, $C = 1\mu F$, $r_a = 2,000$ ohms, $\mathcal{E} = 2,000$ volts. Then

$$V_1 = \frac{2,000}{1 + 2\pi \times 50 \times 2 \times 10^3 \times 1 \times 10^{-6}}$$

$$= \frac{2,000}{1.628} = 1,230 \text{ volts.}$$

During the next positive half-cycle of transformer secondary voltage, anode current will not start to flow until the transformer voltage exceeds 1,230 volts, so that in effect only 2,000—1,230 or 770 volts are available to produce a further charge in the condenser. If this further charge produces an additional counter-E.M.F. of $\frac{770}{1.628}$ volts, and so on during succeeding cycles, the rate at which the condenser P.D. will grow will be shown as below. In this table $\mathcal{V}_1, \mathcal{V}_2, \mathcal{V}_3$, etc., represent the differences between the applied peak E.M.F. and the condenser P.D., the latter quantity for each half-cycle being denoted by V_1, V_2, V_3 , etc.

$$\begin{aligned}
 \mathcal{E} &= 2,000 \\
 V_1 &= \frac{2,000}{1.628} = .615 \times 2,000 = 1,230 \\
 \mathcal{V}_2 &= 2,000 - 1,230 \\
 &= 770 \\
 V_2 &= .615 \times 770 + 1,230 = 472 + 1,230 = 1,702 \\
 \mathcal{V}_3 &= 2,000 - 1,702 \\
 &= 298 \\
 V_3 &= .615 \times 298 + 1,702 = 184 + 1,702 = 1,886 \\
 \mathcal{V}_4 &= 2,000 - 1,886 \\
 &= 114 \\
 V_4 &= .615 \times 114 + 1,886 = 70 + 1,886 = 1,956 \\
 \mathcal{V}_5 &= 2,000 - 1,956 \\
 &= 44 \\
 V_5 &= .615 \times 44 + 1,956 = 27 + 1,956 = 1,983 \\
 \mathcal{V}_6 &= 2,000 - 1,983 \\
 &= 17 \\
 V_6 &= .615 \times 17 + 1,983 = 10.4 + 1,983 = 1,993.4
 \end{aligned}$$

After six cycles the condenser will be charged to a voltage practically equal to the applied peak voltage, and no further action will take place. Though admittedly an approximation, this argument traces out clearly the successive stages by which an applied voltage of sinusoidal form results in the development of a steady P.D. between the condenser terminals. The curve shown in fig. 24 to which reference has already been made, is actually plotted from the data calculated above.

41. Now consider the load resistance to be connected as shown in fig. 23. During the first positive half-cycle the diode will allow an electron current to flow as in the case just discussed, and of this a portion will flow through the resistance and a portion will charge the condenser with the same polarity as before. By the end of the half-cycle the condenser P.D. will reach some finite value V_1 (fig. 25). During the next half-cycle the anode is negative, no anode current will flow, and the condenser will receive no charge. The condenser P.D. will now tend to maintain the current through the resistance R , and will therefore fall, say to V_2 . During the second positive half-cycle an anode current will not be established until the transformer secondary voltage is equal to the condenser P.D., V_2 ; at this instant the reservoir condenser will commence to charge once more, eventually reaching some P.D., V_3 , greater than V_2 , but when the secondary voltage falls below the condenser P.D. no further charge will be introduced. Meanwhile the current through the load resistance has been maintained, and in fact slightly increased, by the combined effects of the transformer voltage and condenser P.D. During the remaining portion of the positive half-cycle and the succeeding negative half-cycle, the condenser receives no charge, but

CHAPTER IX.—PARA. 42

continues to maintain the load current, and the P.D. across the load falls slightly, i.e. from V_3 to V_4 . After a few cycles the energy received by the condenser during the positive half-cycles will be exactly equal to that dissipated in the load resistance during the whole cycle, and the condenser P.D. will be maintained at an average value V_0 , about which it fluctuates slightly as shown in the diagram.

The slight variation in the voltage at the load terminals is called the voltage ripple, and is obviously an undesirable phenomenon. It is convenient to regard it as a small alternating voltage superimposed upon the average condenser voltage V_0 , the fundamental frequency of this ripple voltage being equal to that of the A.C. supply, although higher harmonics are also present.

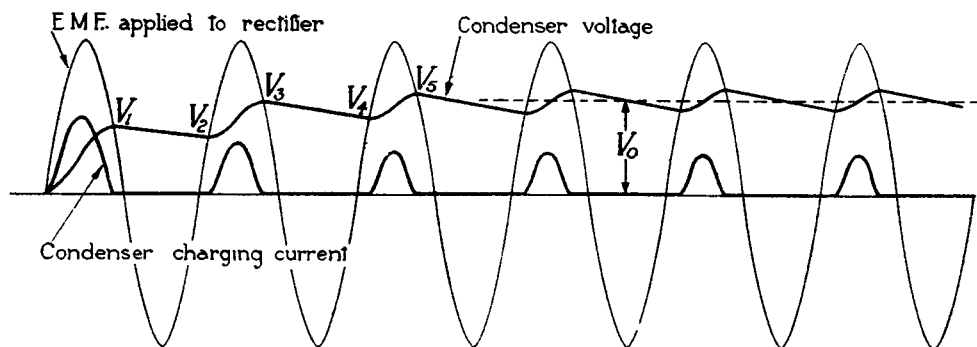


FIG. 25, CHAP. IX.—Charge of reservoir condenser under working conditions (full-wave rectifier).

The effect of the ripple is to modulate the anode current supply to the transmitter, and the resulting radiation is not truly a continuous wave. The percentage variation of the condenser P.D. is, by fig. 25, $\frac{V_5 - V_4}{V_0} \times 100$. This percentage is inversely proportional to the load

resistance, the capacitance of the reservoir condenser, and the ripple frequency. For a given transmitter, fixing the value of the load resistance, therefore, the ripple is less with a high frequency supply voltage than with a low, and it would appear that the reservoir condenser should also be as large as practicable. It must however be observed that the larger the capacitance, the greater is the number of cycles taken to charge the condenser to its final mean value V_0 , and if the capacitance is larger than one or two microfarads the quality of the transmission may become "chirpy", while when signalling at extreme ranges the shorts may possess insufficient power to give an audible response. The supply frequency is usually fixed by considerations other than those of signalling, e.g. in this country, all commercial supply mains will eventually be standardised at 50 cycles per second. It is possible to reduce the percentage of ripple by the use of a full-wave rectifying system.

Full-wave rectifying system

42. (i) In the full-wave rectifying system both half-cycles of the transformer voltage are utilised in charging the reservoir condenser. A typical arrangement is shown in fig. 26. Two separate diodes U_1, U_2 are connected in such a way that they share a common filament supply, but the anode of each valve is connected to one end of the secondary winding of the transformer. The latter carries a tapping at the electrical centre and the reservoir condenser is connected between this point and a point in the rectifier filament circuit, the load resistance being connected in parallel with the reservoir condenser as in the half-wave system. The action of the circuit follows from the previous discussion, and may be outlined as follows. When the switch S is closed, suppose the upper end X of the secondary winding to become positive with respect to

the filaments, and the opposite end Y correspondingly negative. As the voltage rises an electron current is established in the valve U_1 but not in the valve U_2 and electrons flow into the condenser at the plate A which thereby acquires a negative charge, the plate B becoming positive. At the end of the first half-cycle the condenser is charged to the voltage V_1 (fig. 27). During the next half-cycle, the anode of the valve U_2 is positive to the filament, and the electron current through it again charges the reservoir condenser, with the same polarity as before. At the end of one cycle, therefore, the condenser has received two charges instead of only one as in half-

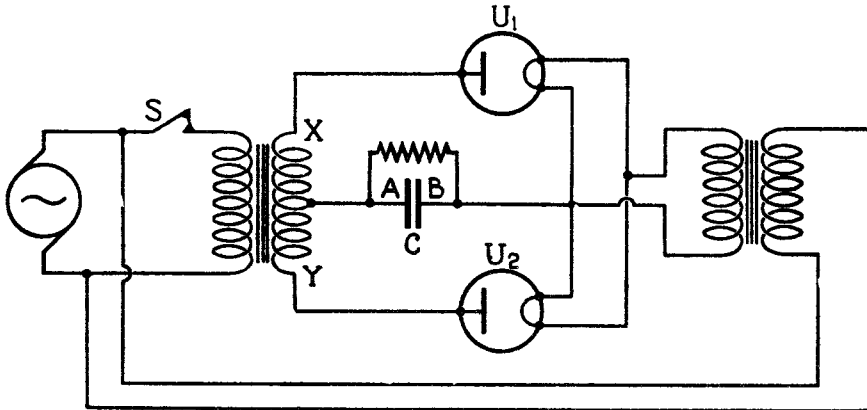


FIG. 26, CHAP. IX.—Full-wave rectifying circuit.

wave rectification. So far as charging is concerned, the method of connecting the diodes has the effect of making both half-cycles of positive sign, and in fig. 27 the negative half-cycles have been inverted to show this effect. The condenser P.D. tends to maintain a current in the load resistance, and in the intervals when the condenser is not actually receiving a charge its terminal P.D. falls; thus the condenser P.D. will increase during successive cycles as shown by the line V_1, V_2, V_3 , etc. During the first few half-cycles, the condenser voltage gradually reaches a state of equilibrium in which the charge given up in order to maintain the load current is just

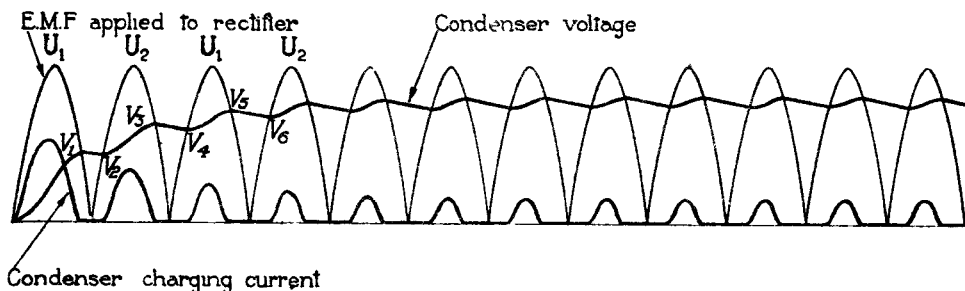


FIG. 27, CHAP. IX.—Charge of reservoir condenser under working conditions (full-wave rectifier).

balanced by the charge received from the rectifying valves. A slight voltage ripple is found to exist, its fundamental frequency being twice the supply frequency. For a given transmitter and reservoir condenser, the percentage ripple in the full-wave system is one-half that of the half-wave system.

(ii) In practice, the wave-form of the transformer secondary voltage is rarely if ever sinusoidal, for even if the wave-form of the alternating supply has this desirable characteristic, the wave-form of the secondary E.M.F. is distorted owing to the presence of a unidirectional current in the

CHAPTER IX.—PARA. 43

secondary winding during certain portions of the cycle, this current being of course that which charges the reservoir condenser. In the half-wave circuit, this current flows always in the same direction through the whole secondary winding, and consequently there is a steady magnetising component of current and a constant flux in the core, upon which the alternating flux is superimposed. The result of this is to increase the time during which anode current flows, and to reduce the peak value of the anode current. The current and voltage waves of an actual half-wave rectifier are shown in fig. 28a, for a circuit having a comparatively large reservoir condenser. These may be compared with corresponding curves for a full-wave rectifier, fig. 28b, which indicate the greater approach to constant load current given by the latter system.

Keying arrangements and safety devices

43. (i) In the service, two types of diode are in general use in rectifying systems for H.T. supply to transmitters. The first is the hard or vacuum type, for example the valve, rectifying V.U.7.A. These are diodes which are exhausted to the highest degree possible, the filament being of hairpin form and the anode consisting of a nickel cylinder. The anode A.C. resistance of the valve V.U.7.A. is of the order of 1,000 ohms. Their rather high internal resistance is a disadvantage of the vacuum type, causing the P.D. at the terminals of the reservoir condenser to fluctuate violently with changes of load, such as would be caused by "keying" the C.W.

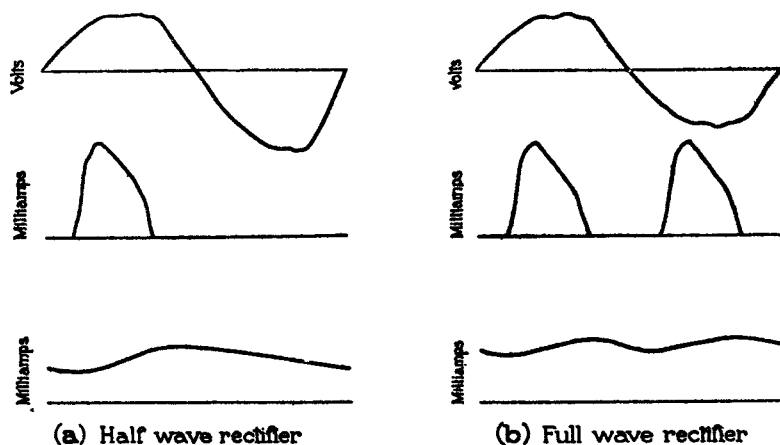


FIG. 28, CHAP. IX.—Secondary E.M.F., charging current, and load current of practical half-wave and full-wave rectifiers.

transmitter. For this reason it is desirable to arrange the keying device in such a manner that it interrupts not only the H.T. supply from the reservoir condenser, but also the primary circuit of the transformer. If this is not done, the condenser P.D. rises to the peak voltage of the transformer winding, in the case of the half-wave rectifier, or to one-half this value with the full-wave rectifier every time the load is entirely removed, and on pressing the key an excessive current may flow in the anode circuits of the transmitter. This is not likely to cause damage, although it may possibly throw undue strain on certain portions of the insulation, nevertheless the excessive voltage is undesirable because it must result in variation of the valve circuit constants and therefore in frequency variation.

(ii) A common method of keying a transmitter supplied by a vacuum-diode rectifying system is shown in fig. 29. The circuits are interrupted in three places, viz., (a) the primary winding of the transformer, (b) the H.T. negative and grid circuits, (c) the H.T. positive supply to the anode of the transmitting valve. This is performed by an electro-magnetic relay key which is operated by a solenoid. The morse key itself is placed in series with the solenoid together with a suitable source of direct current. This diagram also shows two safety devices which are often

incorporated in ground station transmitters. The transmitter is usually mounted in metal panels which totally enclose all those parts which are at dangerously high potential, doors being provided to allow access for adjustment, replacement of valves, etc. Safety switches are then fitted in such a manner that when any door is open the power supply to the transmitter is automatically interrupted. It must be firmly realised that the mere interruption of the power supply does not render all portions of the circuit safe to handle, for the reservoir condenser remains charged to its normal voltage under these conditions. An additional switch is therefore fitted by which the reservoir condenser can be completely discharged. This may be arranged for hand or automatic operation, and is marked "transmit" and "safe" in fig. 29. It can be seen that when the switch is at "safe" the reservoir condenser is short-circuited.

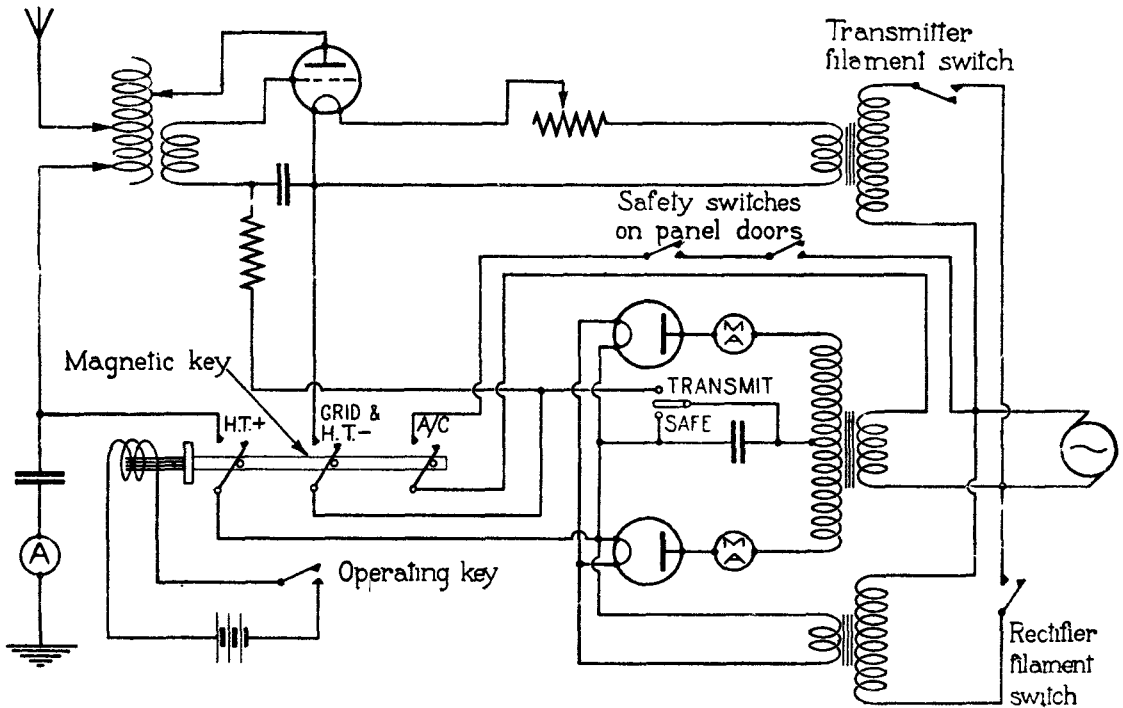


FIG. 29, CHAP. IX.—C.W. transmitter with full-wave rectifying system, showing method of keying and safety arrangements.

(iii) The door switches need not operate directly in the H.T. transformer primary circuit. In modern practice the latter frequently contains an electromagnetic switch similar in principle to the magnetic key, and the door switches are then arranged to break the D.C. supply to the solenoid. This D.C. supply may obviously be obtained from the same source as that from which the magnetic key bobbin is energised, which may be a secondary battery, or may be derived from the A.C. mains through a metal rectifier. The action of this device depends upon the rectification which takes place at a contact between pure copper and a thin film of cuprous oxide. The resistance of such a contact is much higher when the applied E.M.F. is in one direction than in the other, but does not approach infinity as in the vacuum diode. The oxide film will withstand a P.D. of a few volts only, and the practical rectifier consists of a pile of copper discs, oxidised on one side only, which are separated by lead washers and clamped together. Four such units are generally used, the method of connection being known as the full-wave bridge circuit, fig. 30. No reservoir condenser is used as a rule because the existence of an appreciable ripple is not

CHAPTER IX.—PARA. 44

detrimental to their operation as a means of D.C. supply to the magnetic keys and switches above mentioned. Provided that a metal rectifier is not overloaded its life is very much longer than that of a diode rectifier.

The hot cathode mercury vapour diode

44. This rectifier has grown in favour during the past few years for H.T. supply purposes where a moderately high voltage is required. The valve consists of a glass bulb containing an oxide-coated filament and a tungsten anode, and is invariably air-cooled even in the largest sizes. It is easily distinguished by its construction from the vacuum type. The latter is to all intents and purposes merely a transmitting valve with the grid structure omitted, the design of both diodes and triodes of high vacuum being based upon the same considerations. The gas-filled

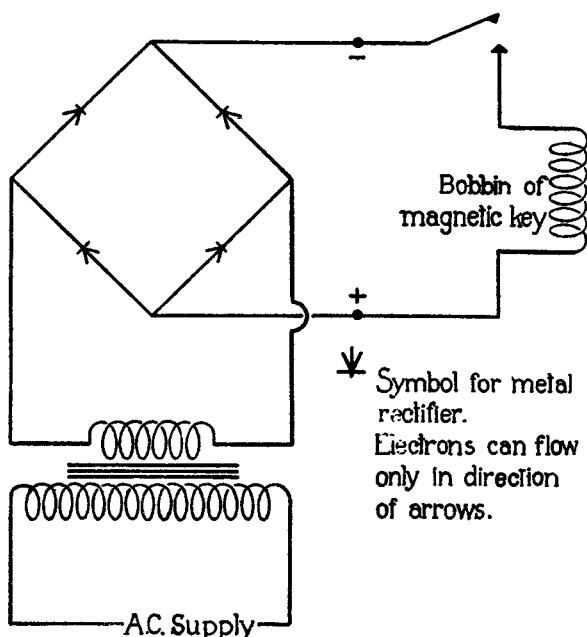


FIG. 30, CHAP. IX.—Full-wave rectifier (bridge circuit) for auxiliary D.C. supplies.

diode is characterised by large inter-electrode spacing compared to the dimensions of the electrodes, the anode in particular being much smaller than the anode of a vacuum diode of equal dissipation. These reduced dimensions are rendered possible by the low power losses in the valve and its consequent high efficiency as a rectifier.

In manufacture the valve is first thoroughly exhausted and a small quantity of liquid mercury is then introduced into the envelope. A proportion of this mercury is vapourised and a state of equilibrium is reached at which the pressure of the mercury vapour is sufficient to prevent further vaporisation, the gas pressure being then of the order of $\cdot 00001$ to $\cdot 00003$ millimetres. The feature of this diode by which it chiefly differs from the high vacuum type is that an anode-filament P.D. of only 15 to 20 volts is sufficient to cause an anode current corresponding to the full electron emission. When the anode is positive with respect to the filament, electrons emitted from the latter encounter mercury vapour molecules, and if the anode-filament P.D. exceeds 10·4 volts the velocity attained by the electrons will be sufficient to cause ionisation

as a result of the collisions. Owing to the small amount of gas present, this ionisation does not lead to any great increase in the anode current, but the positive ions which are formed move so slowly towards the filament compared with the velocity of the electrons in the contrary direction that the negative space charge caused by the latter is practically neutralised, and the anode draws electrons from the filament at the rate at which they are emitted. The anode voltage required to produce the full saturation current is only slightly greater than the potential necessary to cause ionisation, and as the latter is 10·4 volts the full emission is obtainable with an anode filament P.D. of only 15 to 20 volts. Provided however that the anode-filament P.D. does not exceed 22 volts, the mercury vapour ions move so slowly, owing to their comparatively large mass, that no damage is caused to the filament coating by the arrival of positive ions. The important constants of this type of valve are (a) the maximum allowable peak anode current, and (b) the maximum permissible inverse anode voltage, that is the maximum voltage by which the anode may become negative to the filament. The peak anode current is determined by the design of the filament and is not affected by the presence of gas, while the maximum permissible inverse voltage is that which is just insufficient to cause a spark to pass between anode and filament, and is less than the voltage which would cause this effect in vacua. For this reason, the inverse voltage effect is of little importance in high vacuum rectifying valves designed for power supply to transmitters, and was not mentioned in this connection.

45. (i) Owing to the effects of the presence of gas, the mercury vapour diode must be operated under somewhat rigid conditions compared with the high vacuum type. The working temperature must be maintained within certain limits, because the temperature of the valve determines the amount of mercury which will exist in the form of vapour, i.e. if the temperature is too low, more mercury will exist in liquid form and less as vapour, and this will result in a decrease in the number of positive ions formed by collision between electrons and gas molecules. Sufficient anode current can then only be obtained by raising the anode voltage, which will have the effect of producing increased ionisation, but if this increase causes the anode-filament P.D. to exceed 22 volts, the velocity with which the positive ions will impinge upon the cathode will be sufficient to cause disintegration of the latter. On the other hand, an excessive temperature will enable sparking to take place inside the envelope at a lower inverse voltage than normal. A further precaution is necessary, in that an excessive anode current, even if only allowed to persist for a fraction of a second, may cause permanent damage to the valve, for the increased current is accompanied by an increased IR drop in the valve itself and this may exceed the 22 volts above mentioned as the maximum safe anode-filament P.D. The filament must always be brought to full operating temperature before the anode voltage is applied, otherwise during the time taken by the filament to reach the temperature of normal emission, the IR drop in the valve will exceed the safe limit and the filament will suffer heavy damage. For this reason the invariable practice is to fit a device which prevents the application of the anode voltage until some twenty to thirty seconds after the filament circuit has been completed. This apparatus is generally termed the time delay relay, and may be operated by various means. In one type the action is dependent upon the flexure of a bi-metallic strip when heated by a local current, while in another, a coiled spring operates through an escapement similar to that of a clock, the H.T. supply to the anode of the valve being completed only when the specified time has elapsed after the closure of the filament circuit.

(ii) The filament is invariably oxide-coated because it must be designed to function with low terminal voltage, about five volts being the maximum. It is apparent that the P.D. between the ends of the filament must be considerably less than the anode-filament P.D. and the latter rarely exceeds 15 volts. To illustrate this point, assume that the filament P.D. is 8 volts and that the peak voltage of the anode is allowed to rise 15 volts above the most positive part of the filament, then the P.D. between the negative end of the filament and the anode will be 23 volts, and is above the value at which positive ion bombardment of the filament causes disintegration of the latter. This can only be avoided by reducing the anode-filament P.D. to less than 14 volts, a value which is insufficient to obtain full emission from the positive end of the filament. The rule adopted is that the peak value of P.D. between the ends of the filament, plus the peak anode-filament P.D., shall not exceed 22 volts.

CHAPTER IX.—PARA. 46

46. The chief advantages of this type of rectifier are, first its high efficiency, and second its excellent voltage regulation compared with the high vacuum type. The latter feature leads to its adoption for H.T. supply to radio transmitters because it signifies that the voltage variations caused by "keying" the transmitter are comparatively small and do not lead to appreciable frequency variation. The low value of permissible inverse voltage renders it unsuitable for use in normal half-wave or full-wave circuits, and its most usual method of connection is shown in fig. 31 which is called the voltage-doubling circuit. Two similar gas-filled diodes are used, each requiring its own separately-insulated filament supply; it is usual to utilise small step-down transformers for this purpose. One end of the H.T. transformer secondary is connected to one anode and one filament of the valves, while the filament and anode respectively of these valves are connected to the outer terminals of two condensers in series, the centre point being led to the opposite end of the transformer secondary. The action of this circuit is as follows:— Assuming that the first half-cycle of primary current causes the upper end of the H.T. secondary to become positive, an electron current will be established in the valve U_1 , which on reaching the anode will pass through the transformer secondary, charging the plate A of the condenser C_1 negatively, a corresponding repulsion of electrons from the plate B completing the circuit

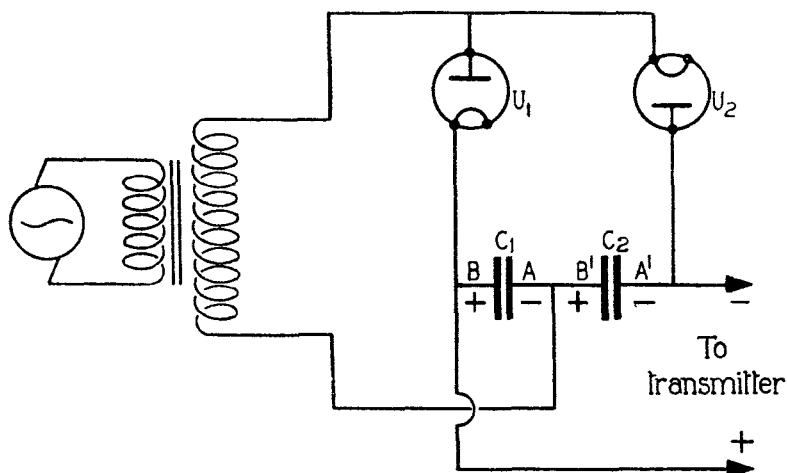


FIG. 31, Chap. IX.—Voltage doubling rectifier circuit for gas-filled diodes.

electron current. During this half-cycle, the valve U_2 is completely inoperative. On the succeeding half-cycle the lower end of the H.T. secondary winding becomes positive, and an electron current is established through the valve U_2 charging the plate A' negatively and repelling electrons from the plate B' leaving it positively charged. The electron current then flows through the transformer secondary to the filament of U_2 . At the end of one complete cycle, therefore, the condensers C_1 and C_2 are both charged, the P.D. between the terminals of each being within a few volts of the peak secondary voltage. The load, i.e. the H.T. circuit of the transmitter, is connected across the outer terminals of the two condensers and is therefore supplied at a voltage which is the sum of the two condenser voltages, or practically double the peak secondary voltage of the transformer, hence the term voltage-doubling circuit.

Let us now consider the inverse voltage to which the valve is subjected. Let the P.D. applied to the load be V_0 and the peak secondary voltage $\mathcal{E}$. Then when the valve U_1 is in operation U_2 is subjected to a peak voltage which is the sum of $\mathcal{E}$ and $\frac{1}{2}V_0$ and as each condenser charges to a voltage practically equal to $\mathcal{E}$ the inverse voltage is very nearly equal to V_0 . This may be compared with the inverse voltage to which the valves are subjected in the full-wave rectifying system. Here the P.D. across the load is also V_0 and is nearly equal to the half-secondary voltage, $\mathcal{E}$. The valve which is not passing an electron current has a peak inverse

voltage which is equal to the half-secondary voltage plus the P.D. across the reservoir condenser, which is greater than $2V_0$. For a given value of V_0 , therefore, the voltage doubling circuit is preferable to the full-wave system if gas-filled valves are to be employed.

FREQUENCY CONTROL

Frequency variation

47. (i) The development of the valve transmitter in the forms already described was such a great advance upon previous methods of generating high-frequency oscillations that these have been almost entirely supplanted. Its introduction led to an enormous increase in the number of radio transmitting stations throughout the world, and with the growth of radio communication a new problem arose, i.e. the question of interference at the receiver by stations other than those with which communication is required. Much can be done to minimise this interference by careful design and operation of the receiver, and this aspect of the subject will be taken up in later chapters, but for the present it must be realised that full advantage of the selectivity of any receiver can only be obtained if certain requirements are fulfilled by the transmitter. These are (a) the frequency of the wave emitted must remain as nearly constant as possible during the whole course of any transmission or series of transmissions, otherwise the receiving operator must keep his apparatus in a process of continual re-adjustment, which calls for a high degree of skill and leads to the necessity for repetitions, etc., even with a capable operator. (b) When for a given frequency a certain set of adjustments have been found and standardised, this frequency shall be emitted whenever the particular adjustments are resorted to, even if a considerable time has elapsed since the original calibration took place. If this condition is not satisfied it is fruitless to make precise frequency allocations according to power and geographical distribution of stations with a view to the reduction of interference.

(ii) The closeness of frequency allocation depends upon the nature of the wave emitted. Every transmitter radiates, not a single frequency, but a band of frequencies, although this may appear to be at variance with the preceding text. It is quite true to say that a C.W. oscillator may emit a single frequency, but intelligence can only be transmitted by breaking up the transmission in some way, e.g. by telegraphic key in the case of the C.W. transmitter. The mere fact that the wave is interrupted causes additional frequencies to be introduced; these are in the immediate neighbourhood of the nominal frequency and must not be confused with the harmonics previously mentioned. The C.W. transmitter operating upon a given frequency radiates power in a band of frequencies covering about 250 cycles above and below the nominal frequency, while the total band occupied by an I.C.W. transmitter may be as much as 10 k.c/s.

48. In transmitters using direct aerial excitation, the principal cause of frequency variation is the change of aerial capacitance which occurs whenever the aerial system moves with reference to its counterpoise or earth system, and in many instances this effect completely masks all other causes. It was therefore the first point to be attacked, the remedy being to use some form of indirect coupling between the actual valve output circuit (which then becomes a closed oscillator) and the aerial circuit itself. Any form of inductive or capacitive coupling may be employed, the choice being made chiefly from the aspect of ease of manipulation, although certain forms are better than others from the point of view of freedom from harmonic radiation. In order to obtain maximum transfer of energy to the aerial circuit, both the coupled circuits must be tuned to the same frequency, or at any rate very nearly so. An increase in the degree of coupling beyond the critical value at which the resonance curve ceases to preserve a single peak will not result in an increase in aerial current, but if this coupling is exceeded a curious effect may occur. The oscillatory circuits as a whole have two resonant frequencies and the valve tends to maintain oscillations at that frequency for which its effective load resistance is least. As the two frequencies are close together and the effective resistance at either frequency is not much different from the resistance at the other, any small alteration in effective resistance may cause a sudden change

of frequency, the effect being called "frequency jump." To avoid this a slight mistuning of the aerial circuit is sometimes necessary in spite of the reduction of power transferred to the aerial circuit. Even when using a closed oscillatory circuit, however, the frequency generated by a triode does not remain perfectly constant. Two principal reasons may be ascribed, viz. (i) the dependence of the frequency upon the valve constants. (ii) The variation in value of the circuit properties, i.e. the inductance, capacitance and resistance. Hitherto it has been assumed that the frequency is controlled entirely by the inductance L and capacitance C of the oscillatory circuit, being given

by the equation $f = \frac{1}{2\pi\sqrt{LC}}$ Even under the simplest conditions, i.e. sinusoidal operation

excluding grid current, and neglecting the effect of inter-electrode capacitance, this equation is

only an approximation, the frequency generated being given by $f = \frac{1}{2\pi} \sqrt{\frac{1}{LC} \left(1 + \frac{R}{r_a} \right)}$ where

R is the effective resistance and is supposed to be "lumped" in the inductive portion of the circuit. The frequency therefore depends upon the anode A.C. resistance of the valve, and this in turn depends upon the operating conditions, for the valve characteristic is never perfectly straight and r_a varies slightly with change of filament current and of anode voltage. Every variation in either of the supply voltages therefore causes some slight change of frequency. The variation of circuit properties is chiefly caused by the mechanical expansion of components as their temperature increases owing to the heat developed. Much ingenuity has been applied to the problem of making these changes self-compensating, for example, by so arranging the inductance that its increase of dimensions causes a corresponding variation of the capacitance and thus maintaining a constant value for the product LC , but such devices are not at present used in the service. A practical solution is to maintain the temperature of the compartment containing the essential components constant and above ordinary room temperature, by some form of thermostat.

Master oscillator system

49. A partial solution of the problem of frequency variation can be achieved by generating only a low power oscillation, the frequency of which is maintained as constant as possible in the particular circumstances. It is possible to design a low power oscillator maintaining a high standard of constancy, because the heat developed, and therefore the temperature rise, is much less, while as the load on the H.T. generator is comparatively small its regulation is good. The

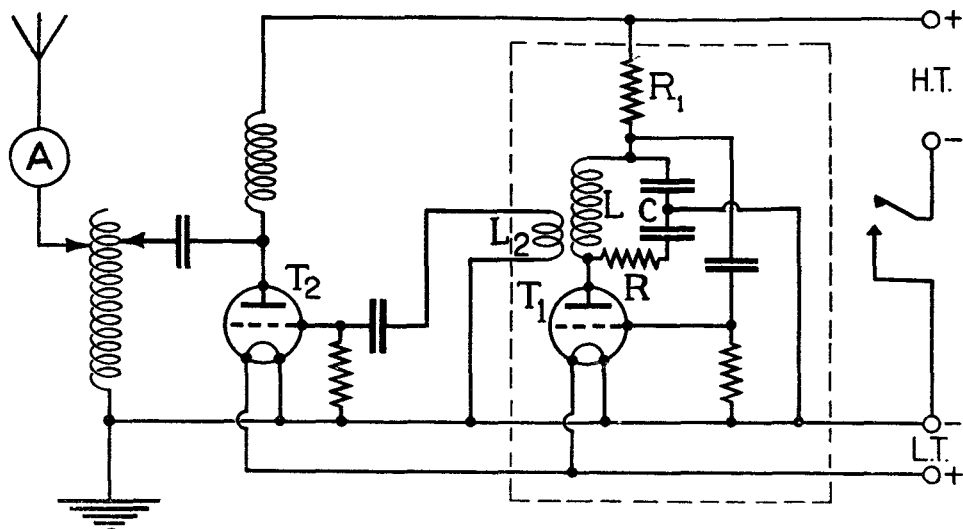


FIG. 32, CHAP. IX.—Simple valve master-oscillator-controlled transmitter.

low power oscillation is then caused to supply grid excitation to a power amplifier, consisting of one or more stages of radio-frequency amplification. The output power of this amplifier is delivered to the aerial circuit, the frequency of the radiation being that of the original low power or master oscillator; under correct operating conditions, the tuned circuits incorporated in the power amplifier have no influence upon the radiated frequency.

A circuit showing the essential features of a master-oscillator-controlled transmitter is given in fig. 32, in which the apparatus shown inside the dotted rectangle forms the master-oscillator. The latter thus consists of a triode valve T_1 , so arranged that oscillations are generated in the anode circuit, the particular form of oscillator in this instance being the Colpitts circuit. The oscillatory flux about the anode inductance L links with the coupling coil L_2 and induces in it an E.M.F. of the generated frequency, which is applied between grid and filament of the power amplifier valve T_2 . The resulting changes of anode current cause oscillatory power to be generated in the aerial circuit at the frequency of the master-oscillator. For maximum output from the power amplifier, its anode circuit must be tuned to the frequency which it is desired to amplify, i.e. the aerial circuit and the closed circuit of the master-oscillator must be adjusted to the same frequency.

Self-oscillation in power amplifier

50. Referring to fig. 32, let us suppose that the oscillator valve T_1 is removed, but the circuit otherwise remains the same. Every component has some slight capacitance to every other, and to earth, while it is possible that many of the connecting leads will have some slight mutual induction with others, no matter how carefully they are spaced. Under certain conditions, the aerial circuit of the power amplifier may be resonant with the grid circuit consisting of the coupling coil L_2 and the stray capacitance across its ends. Now these two resonant circuits are coupled together by the inter-electrode (anode-grid) capacitance of the amplifier valve, and will therefore act as a tuned-anode/tuned-grid oscillator. Hence we find that at this particular frequency the power amplifier itself acts as a transmitter and the master-oscillator is redundant so far as the production of aerial oscillations are concerned. The frequency emitted however is now subject to all the causes of variation already discussed, and the possibility of the transmitter operating in this way, even if the oscillator valve is returned to its normal position, is obviously undesirable.

Neutralisation of power amplifier

51. The steps necessary to remove all risk of the power amplifier valve acting as an oscillation generator are easily seen. Since the oscillations are produced owing to the existence of coupling between the anode and grid circuits, either this coupling must be destroyed, or a coupling of equal magnitude and opposite sign introduced between the two circuits. A possible solution is as follows. Since for oscillation to be maintained by magnetic coupling, the sign of the mutual inductance between anode and grid coils must be negative, the tendency to oscillation owing to the anode-grid capacitance coupling may be annulled by a certain amount of positive mutual inductance between anode and grid coils, as shown in fig. 33a. The principal disadvantage of this solution is the practical one of designing the mutual inductive coupling in such a way that it will be sufficiently great at the lowest frequencies, and yet be capable of extremely fine adjustment so as to counterbalance exactly the capacitance coupling.

A simple modification of this method is found very effective in practice. Instead of coupling the output and input (i.e. anode and grid) circuits together, an additional coil is loosely coupled by mutual induction to the anode coil, and is connected between grid and filament, a very small variable condenser being interposed, as in fig. 33b. The power amplifier valve thus virtually possesses two input circuits, one from the master oscillator, and the other coupled to its own output. The sign of the mutual induction in the latter coupling is such as to oppose the maintenance of oscillations, and the actual voltage applied between grid and filament by this coil is controlled by varying the capacitance of the small variable condenser. By this means any tendency to self-oscillation can be completely neutralised, and the method is eminently suitable for use in aircraft transmitters, being robust and easily adjusted.

CHAPTER IX.—PARAS. 52-53

Master-oscillator-controlled transmitters for aircraft

52. The essential features of this type of transmitter can now be summarised. The primary necessity is that the frequency of the oscillator itself shall be as constant as possible, bearing in mind the conditions under which the transmitter is to operate. A ground station, operating upon a fixed frequency, can be maintained within about $\cdot 01$ per cent. of that allotted. This high standard is only possible if with the aid of numerous subsidiary devices, mention of which will be made in due course. At the other extreme is the aircraft transmitter in which space and weight are primary considerations in design, while robustness and freedom from breakdown are of equal importance once the instrument is accepted for service use. Again, a given aircraft may be employed in many different duties at various times, thus necessitating radio operation over an extremely broad frequency spectrum, and the transmitter must be capable of operation over the whole service range of frequencies, otherwise it must be changed when necessary to suit the particular operation upon which the aircraft is to be engaged. Considerations of space, weight and time involved in change of apparatus all indicate that the ideal aircraft transmitter should cover any frequency upon which the aircraft may operate, and it is unreasonable to expect a

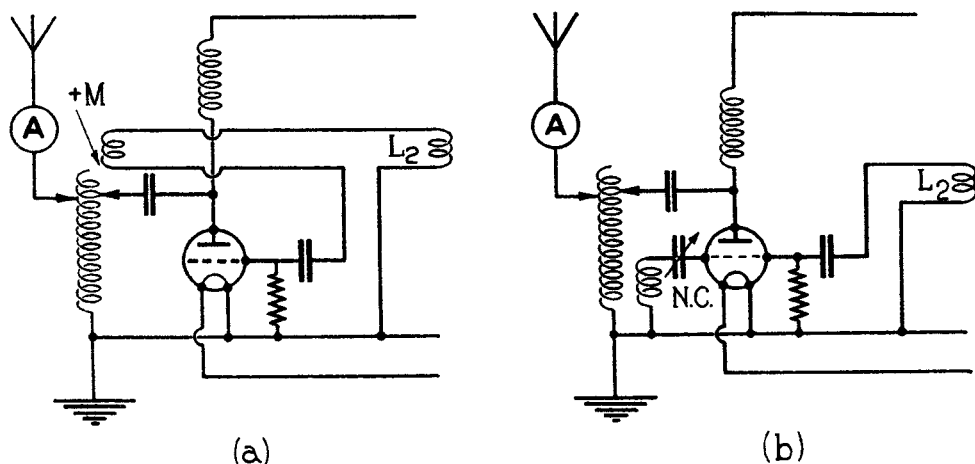


FIG. 33, CHAP. IX.—Theoretical and practical forms of neutralisation of amplifier circuit.

frequency constancy even approaching that achieved under the ideal conditions of single frequency operation on the ground. In aircraft transmitters therefore, the master-oscillator system must be designed to give the highest constancy obtainable without the aid of subsidiary devices such as temperature control, and every effort is made to make the master-oscillator itself a constant frequency source.

Constancy of frequency of master-oscillator

53. In a previous paragraph it was pointed out that the frequency generated by the master oscillator would depend, even in the simplest instances, upon the effective resistance of the tuned circuit and upon the anode A.C. resistance of the oscillator valve, as well as upon the effective inductance and capacitance, including in the latter the effects due to mutual inductance and stray capacitance. Now the effective resistance of the tuned circuit is not actually lumped in the circuit at one point, but depends upon a number of power-dissipating factors, such as the power transferred to the grid circuit of the power amplifier, eddy currents in neighbouring metals, dielectric losses in surrounding insulating material, and losses due to the flow of grid current. Ignoring for the present the variations in effective inductance and capacitance, consider the anode A.C. resistance and grid A.C. resistance, which are denoted by r_a and r_g respectively. These are to some extent dependent upon each other, and upon the amplitude of the generated oscillations,

thus, for a fixed value of grid bias, an increase of amplitude leads to increased grid current, and the slope of the $I_g - V_g$ curve increases rapidly for positive values of grid voltage. Hence the increase of amplitude is accomplished by an increase of grid A.C. conductance, or a decrease of grid A.C. resistance. On the other hand, provided saturation value of anode current is not attained, this increase of amplitude will affect r_a very little. Now consider the conditions when grid bias is obtained by the condenser and leak method. The greater the amplitude, the more negative the bias, and the mean anode current decreases, the working point finally reached being near the curved portion at the foot of the $I_a - V_g$ curve, hence the anode A.C. resistance of the valve increases with amplitude of oscillation. As the final value of bias taken up by the grid is that at which a small steady grid current flows, the grid A.C. resistance corresponds to this value in any case and is not greatly affected by the amplitude. The final amplitude of oscillation corresponds to some particular values of r_g and r_a such that the power generated by the oscillator valve is exactly equal to the total power dissipated in the oscillator, including all the circuit losses, which are expressed by the effective resistance R . Efforts to design a constant frequency oscillator are therefore in the direction of minimising the variations in r_a and r_g , or by arranging the circuit in such a manner that the frequency generated is independent of these resistances. The first step is to arrange that the supply voltages are maintained as constant as possible, and this problem presents more difficulty than is apparent on the surface, particularly in aircraft. For L.T. supply a battery is preferable to an air-driven generator, and baretters may be inserted in series with the filaments. The baretter consists of a resistance of iron wire or some alloy, which is enclosed in a bulb which is filled with hydrogen. The resistance of the element is not constant, but depends upon the current flowing, i.e. the iron has a large positive temperature coefficient. The use of such a baretter thus tends to maintain the current at a constant value, but its utility is limited by the fact that it is not instantaneous in action.

54. In aircraft, the most convenient method of obtaining H.T. supply for a complete master-oscillator-controlled transmitter is by the employment of a generator, capable of supplying up to about 2,000 volts. The regulation of this generator is of considerable importance, for the load upon it is constantly varying during every cycle of oscillation, because the value of the anode A.C. resistance of the valve is not constant during the whole of the cycle. The frequency variation due to this can be reduced by supplying the anode circuit of the master-oscillator through a swamping resistance, of a value equal to or slightly greater than the valve A.C. resistance. The presence of this resistance (R_1 in fig. 32) necessitates an H.T. voltage about double the working anode-filament voltage of the oscillator valve, but this is not a serious matter as the power-handling capacity of this valve is much less than that of the amplifier valve, and a higher value of H.T. voltage would be required for the latter in any circumstances. The oscillator circuit must be designed with the highest magnification practicable, and in order to achieve this the total effective resistance of the oscillatory circuit must be as low as possible. As the power supplied to the amplifier circuit represents a portion of the effective resistance of the oscillatory circuit, this power must also be small. Next, the effective r_a must be as high as possible, so that variations of this quantity make little change in the ratio $\frac{R}{r_a}$. This is attained by operating with considerable grid bias obtained by the leaky condenser method, which enables a large value of r_g to be maintained simultaneously, and thus tends to give higher circuit magnification. From this point of view a high-resistance grid leak is indicated.

Methods of obtaining frequency constancy by designing circuits in which the frequency generated is independent of r_a and r_g have been little used in this country, although they have found some favour in America. They depend upon the use of phase-shifting circuits between the valve and the oscillatory circuit, and the additional complexity renders them unsuitable for use in transmitters subject to frequent alteration of operating frequency.

Tuning a master-oscillator-controlled transmitter

55. The general principles outlined in paras. 14 and 15 apply to the tuning of all installations, but where a master-oscillator is fitted it is not possible to couple the wavemeter to the oscillatory circuit and so adjust the latter to the desired frequency, owing to the nearly perfect screening

CHAPTER IX.—PARA. 56

which is essential for satisfactory operation. As a rule, therefore, the oscillatory circuit is approximately calibrated, at any rate at a few points, and the procedure of tuning may be outlined as follows:—The master-oscillator is set to the desired frequency as nearly as possible by interpolation between the points of calibration, and is switched on, the capacitance of the neutralising condenser for the amplifier valve being first set to zero. Switching arrangements usually provide that for tuning, reduced power is applied to the oscillator valve, while the amplifier valve is completely inoperative. Owing to the un-neutralised anode-grid capacitance of this valve, however, a small oscillatory voltage will be developed in the aerial circuit, and the resulting current may be observed by means of a suitable low-reading thermo-ammeter. The aerial circuit is now brought into resonance with the master-oscillator by adjusting the aerial tuning inductance until the aerial current reaches a maximum value. Before applying full power to the transmitter, the anode-grid capacitance of the amplifier valve must be neutralised, by adjustment of the neutralising condenser, until the aerial current is reduced to zero, and when this operation has been performed, the low-reading thermo-ammeter is switched out of the aerial circuit and a meter of full current-carrying capacity substituted. By means of suitable switches, the amplifier valve is rendered operative, and up to about two-thirds of the maximum power input supplied to the transmitter, the frequency radiated by the aerial circuit being then measured in the usual manner. A slight re-adjustment of the oscillator frequency may be made without the necessity for a repetition of the neutralising procedure, but if the frequency differs considerably from that anticipated, it is necessary to switch off, re-set the master-oscillator to a slightly higher or lower frequency, and repeat the neutralisation. In all cases, when dealing with service transmitters, the tuning instructions given in A.P. 1186, Signal Manual, Part IV, should be strictly adhered to. In order that an aircraft transmitter may be set up to the desired frequency on the ground, it is necessary to simulate the effect of the aircraft aerial system by a suitable circuit in order that the aerial tuning inductance may be adjusted to an approximately correct value, and the neutralisation correctly performed. The apparatus used for this purpose is known as an artificial aerial, and may consist of a fixed inductance of a few microhenries, a variable condenser, and a series resistance having a value approximately equal to the radiation resistance of the aerial system to be used. The capacitance may be of the order of from $\cdot 00002 \mu F$, simulating that of a fixed aerial for use in the 10 to 20 M.c/s band, to about $\cdot 0003 \mu F$, representing the capacitance of a trailing aerial of about 250 feet. Instructions relating to each particular design of artificial aerial will be found in the tuning instructions referred to above.

Frequency control of master-oscillator

56. Reference has already been made to the possibility of attaining an enhanced degree of frequency constancy in the master-oscillator itself by means of auxiliary apparatus, in circumstances which permit the increase of space and weight and the additional complexity entailed. These measures may be divided into two classes, first, thermostatic control of the temperature of the master-oscillator and its associated circuits. Second, the employment of some form of electro-mechanical oscillator having a high degree of frequency constancy as a control for the electrical frequency. Thermostatic control entails the location of the whole of the master oscillator, including the valve itself, in a compartment which is automatically maintained at a temperature slightly higher than would be reached by the apparatus in the absence of temperature control. This compartment is heated by an electric radiator, and the amount of current flowing in the heating circuit is controlled by an instrument called a thermostat. The most sensitive thermostat is the toluene type, the principle of which is as follows. A vessel having a large surface area contains a quantity of liquid toluene, which has the property of expanding considerably when subjected to only a slight increase of temperature, that is, its coefficient of expansion is very high. A glass U-tube connected to this vessel contains a quantity of mercury and in one vertical member the mercury level rises and falls with the variation in the volume of toluene, and therefore with the temperature. A platinum wire is fused through the tube well below the normal mercury level, and a similar wire at a point just above the normal level. When the mercury rises to the level of the latter wire, the two are bridged conductively, and complete the operating circuit of a relay, the local contacts of which then break the electric supply

to the radiator. The temperature of the compartment falls until the contraction of the toluene allows the mercury level to fall clear of the upper wire, and when this occurs the relay contacts close, again completing the circuit through the radiator. By careful design a toluene thermostat can be made to maintain the temperature of the compartment constant to within $\cdot 01^{\circ}\text{C.}$, and by special construction it is claimed that a constancy within $\cdot 001^{\circ}\text{C.}$ has been achieved. A simpler form of thermostat is also in use, in which the variation of temperature causes flexure of a bi-metallic strip, the latter carrying a contact which opens or closes the heating circuit. This device is not so sensitive as the toluene thermostat but is comparatively simple and robust and occupies little space.

Three forms of electro-mechanical oscillator are used for frequency control, viz.

- (i) the electrically driven tuning fork,
- (ii) the piezo-electric crystal.
- (iii) the magnetostriction oscillator.

Only a brief outline of the principles of the first two can be given here. The third method is only rarely adopted.

Tuning fork control

57. The tuning fork has long been recognised as a standard of frequency, forks for the tuning of musical instruments being generally designed to vibrate at a frequency corresponding to the note A next above middle C (440 cycles per second). Provided the temperature is constant and the amplitude of vibration low, the tuning fork maintains its nominal frequency with extreme accuracy. For frequency control, it is usual to employ a fork having a frequency lying between one and two k.c/s. This is very much lower than the frequency of transmission and it is necessary to have recourse to considerable amplification, the frequency being doubled or trebled at each stage.

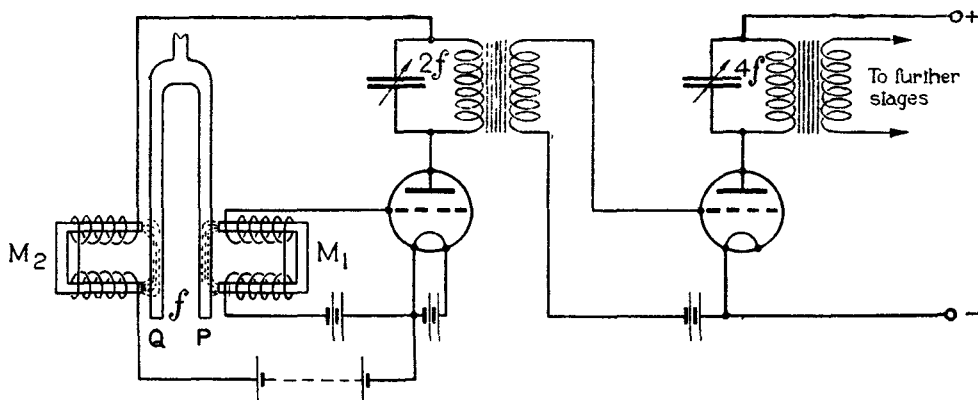


FIG. 34, CHAP. IX.—Frequency control by tuning fork.

The number of such stages is reduced by the employment of a high initial frequency, but unfortunately it is difficult to ensure satisfactory maintenance by electrical means at frequencies higher than about 2 k.c/s. The fork is maintained in vibration at its natural frequency by the arrangement shown in fig. 34. Near the base of each prong is mounted an electromagnet, the coils of which are wound upon soft iron pole pieces which are extensions of the poles of a permanent magnet as in the telephone receiver; in practice a pair of telephone receivers suitably modified may actually be utilised for the purpose. The coils of one magnet are connected in the anode circuit of a triode valve, and the coils of the magnet operating upon the other prong are connected between grid and filament, suitable grid bias being arranged. The fork is analogous to an electrical circuit having very small damping, and therefore very little power losses. This equivalent circuit may be considered to be electromagnetically coupled to both grid and anode

CHAPTER IX.—PARA. 58

circuits, and once set in oscillation can be maintained in this state by the conversion of a small amount of electrical energy into mechanical energy during each cycle of oscillation. Vibration can be initiated by a slight tap on the base of the fork, and the subsequent action is somewhat as follows. At a given instant suppose both the prongs P and Q are moving outwards, then the reluctance of each magnetic circuit is decreased and the flux through each coil increases with a consequent induced E.M.F. in each coil. The E.M.F. induced in the coil M_1 is applied between grid and filament of the valve causing a change of anode current, and if this variation causes an increase in the flux density of the magnet M_2 , the prong Q will be attracted more strongly than by the normal flux. As the prongs reach the limit of their outward excursion the induced E.M.F. in the coil M_1 falls to zero, and when owing to their elasticity the prongs move inward, the increase of reluctance in the air gap between M_1 and P results in the production of a further induced E.M.F. which is applied between grid and filament of the valve. The consequent change of anode current in turn causes a reduction of the pull of M_2 upon its prong, and the latter therefore moves inward to a greater distance than would otherwise be the case. In this way the fork is maintained in vibration, and adjustment of the frequency within small limits can be achieved by loading the ends of the prongs with metal washers. The mean grid potential is maintained at such a value that the anode current changes are unsymmetrical, and a circuit tuned to the second harmonic of the fork frequency is included in the anode circuit of the valve. The second harmonic frequency is selected by this circuit for magnification, and this frequency is further amplified with a similar frequency doubling process at each stage until the desired frequency is reached, after which power amplification is employed in order to supply the aerial with the necessary power. If the frequency of the fork is 1 k.c/s for example, eight frequency doubling stages will give an output at 256 k.c/s. It is obvious that this method of operation in conjunction with thermostatic control of the temperature of the fork itself will give a high degree of constancy, probably being superior to all other methods, but it suffers from the following limitations, (i) its use is limited to certain exact multiples of the tuning fork frequency, i.e. those obtainable by frequency multiplication, (ii) the large amount of auxiliary apparatus can only be justified in stations of comparatively large power. Certain B.B.C. broadcasting stations are controlled in this manner.

Crystal control

58. Crystal control is suitable for use in the higher frequency bands, i.e. above about 1,000 k.c/s at which tuning fork control would become impracticable owing to the large number of frequency-multiplication stages which would become necessary. Certain crystalline substances, notably quartz, tourmaline and rochelle salt, possess the property to which the name piezo-electric (from the Greek piezo, to press) is given. The phenomena associated with this property are briefly as follows. When a suitable plate of the material is subjected to mechanical stress, electric polarity is developed upon opposite faces of the plate. A reversal of the nature of the stress, for example a change from tension to compression, causes a reversal of the electric polarity. Conversely, if a difference of electric potential exists between two opposite faces, mechanical strain or deformation of the crystal occurs, and the sense of this strain is reversed by a reversal of electric polarisation. The material in almost universal use for frequency control practice is quartz which is comparatively cheap and robust. Tourmaline is used in certain instances, but rochelle salt (sodium-potassium tartrate) which possesses the piezo-electric property to a much higher degree than any other substance, is rarely used, owing to its mechanical and electrical weakness. Quartz is a mineral which is found in crystalline form chiefly in Madagascar and Brazil. It is an oxide of silicon (SiO_2) and the silica used in the construction of large power transmitting valves is actually fused quartz. The natural quartz crystal, fig. 35, is of an irregular hexagonal section, opposite sides being parallel, and the ends are pointed, taking a form which is nearly pyramidal. The axis passing from end to end along the length of the crystal is called the Z or optical axis, because a section cut from the crystal perpendicular to this axis exhibits certain optical properties which need not be considered. Two other systems of axes may also be drawn, namely the three axes X_1 , X_2 , X_3 , passing through opposite corners of the section, which are called electrical axes, and the system consisting of the three axes Y_1 , Y_2 , Y_3 , which are perpendicular

to the faces of the crystal. These are called the mechanical axes. A slab cut from the crystal with flat sides perpendicular to an electrical axis, fig. 36a is called an X (or Curie) cut while a slab cut with flat sides perpendicular to a mechanical axis, fig. 36b, is called a Y (or 30°) cut. For different applications of the piezo-electric effect, either of these "cuts" may be preferable. A distinctive property of such crystal is that it possesses a well marked mechanical resonant frequency, which depends upon the density and the elasticity of the quartz. The natural frequency of a slab of quartz is found to depend upon the axes in which the cut was made, and upon the thickness, of the slab, thus an X-cut slab has its "thickness" in the X direction, and its natural frequency is found by experiment to be approximately $\frac{2.86 \times 10^8}{t}$ k.c/s. while a Y-cut slab has a natural frequency approximately equal to $\frac{1.96 \times 10^8}{t}$ k.c/s., t being the thickness in millimetres.

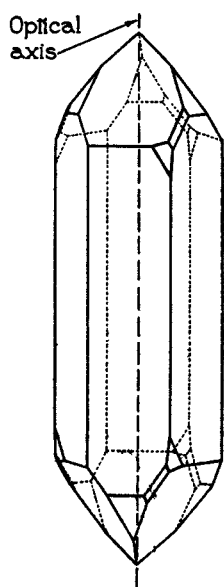
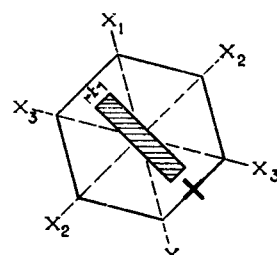
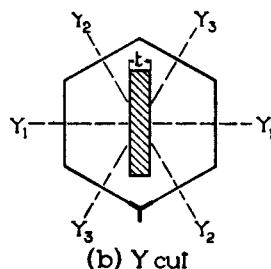


FIG. 35, CHAP. IX.
Natural quartz crystal.



(a) X (or Curie) cut



(b) Y cut

FIG. 36, CHAP. IX.—Axes of quartz crystal shewing X and Y cut plates.

59. (i) The electrical properties of a quartz crystal may be shown by mounting it between two metal plates so that it constitutes an electrical condenser of small capacitance and connecting it in parallel with the tuning condenser of a closed oscillatory circuit, which is then adjusted to a frequency near to the known or estimated natural frequency of the crystal. The closed circuit is then coupled to a calibrated valve oscillator which is capable of very fine adjustment of frequency. The potential difference between the crystal faces may be measured by a suitable radio-frequency valve voltmeter, and it will be observed that as the oscillator frequency is varied the voltmeter reading gives the resonance curve of the closed circuit, but instead of being perfectly smooth over its whole range, there is a steep and sharp voltage drop over a range of a few cycles, the optimum drop occurring at the natural frequency of the crystal, and a typical curve showing the effect is given in fig. 37. The portion of the curve marked A B C is referred to as a "crevasse," and the cause of a crevasse is the absorption of power by the quartz crystal, this power being expended in producing mechanical oscillations. In practice, owing to various causes, a crystal may have several natural frequencies spaced very closely in the spectrum, and this is obviously undesirable. Even if the quartz itself is of a suitable quality, multiple crevasses may be caused

CHAPTER IX.—PARA. 59

by lack of parallelism between the two faces of the crystal, or by minute flaws or scratches in the material. The high cost of really good quartz crystals is due to the tedious process of cutting and grinding which may prove to be wasted if the crystal cannot be so manipulated as to give one clearly defined natural frequency well separated from any subsidiary resonances which may exist. The equivalent electrical circuit of the quartz crystal in its mounting is shown in fig. 38, in which C_1 represents the capacitance of the small condenser formed by the quartz dielectric and metal electrodes while the inductance L capacitance C and resistance R represent electrical analogues of the mass, stiffness and mechanical resistance or dynamic friction. The interest of

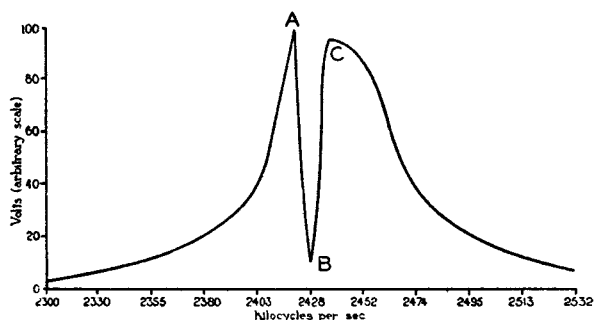


FIG. 37, CHAP. IX.—Resonance curve of quartz crystal shewing crevasse.

this method of representation lies in the values which must be allotted to the constants L , C and R , and may be exemplified by a certain crystal for which these values were accurately determined. The natural frequency of the crystal in this instance was about 44 k.c/s, the object of choosing such a low frequency crystal being to avoid errors of measurement, which is difficult at higher frequencies. The equivalent inductance was found to be 160 henries and the capacitance $\cdot 08$ micro-microfarad, while the equivalent resistance was 1,500 ohms. The magnification of the equivalent electrical circuit is therefore approximately 30,000, whereas in an actual electric

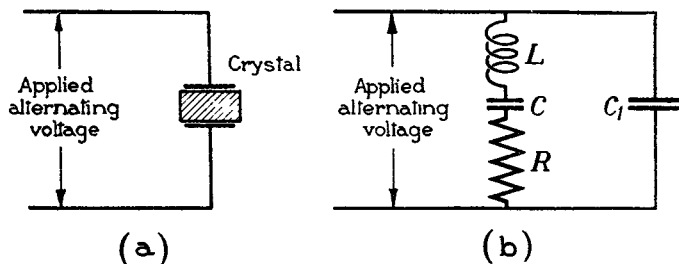


FIG. 38, CHAP. IX.—Quartz crystal under applied alternating voltage, and equivalent electrical circuit.

circuit tuned to this frequency it would be difficult to achieve a magnification of 1,000. In fig. 39 the crevasse is shown upon a very open frequency scale, and it will be observed that a forty per cent. variation of oscillatory voltage occurs within a frequency of about 2 cycles each side of the crystal frequency.

(ii) The initial selection of natural quartz crystals likely to produce suitable oscillators is performed by inspection, the method of natural formation of the crystal being a rough guide. Trial slabs cut from the crystal are then ground flat and parallel, saws fed with diamond dust being used for cutting, and carborundum of various grades for rough grinding. When the latter

stage is reached the slabs are tested for piezo-electric properties by the application of mechanical stress, the faces of the slab being placed between electrodes which are connected to an electrometer. The finished crystal usually takes the form of a disc of about 1 inch diameter or a square of approximately 1 inch side, the thickness depending upon the natural frequency required. Final

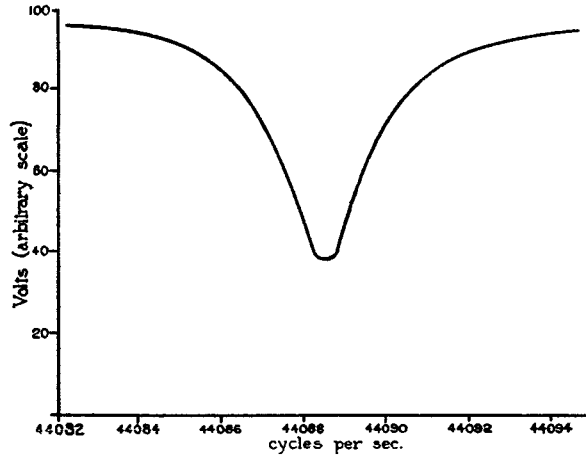


FIG. 39, CHAP. IX.—Crevasse in resonance curve of quartz crystal.

grinding to fine limits is performed upon an optically flat surface with rouge, and the frequency determined by actual trial in a suitable oscillatory circuit. It is important that the crystal should be allowed to attain its working temperature before deciding that further grinding is desirable.

Crystal mountings

60. Three types of mounting have been developed and all are used to some extent. Least frequently employed is that in which the crystal itself has a metallic electrode sputtered directly upon each face, the crystal being then mounted on its edge between light springs, or laid hori-

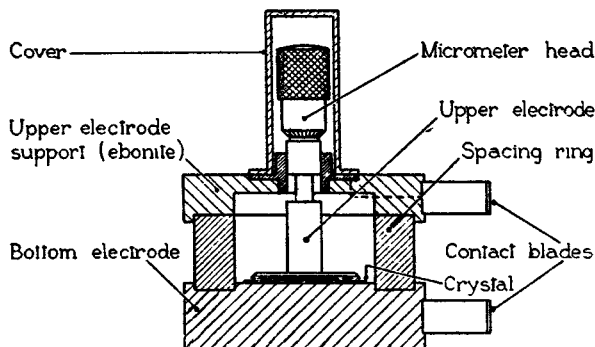


FIG. 40, CHAP. IX.—Mounting for quartz crystal.

zontally upon a conducting surface, a copper strip contact resting upon the upper metallised face of the crystal. A more common form used with unmetallised crystals is shewn in fig. 40. It consists of a dust-tight case, which contains a lower electrode upon which the crystal lies in a shallow circular trough. The upper electrode is mounted upon a micrometer screw by means of

CHAPTER IX.—PARA. 61

which it can be brought as near to the crystal as may be desired, without actually touching it. The employment of an air gap is desirable for two reasons. First, the mechanical load upon the vibrating crystal is reduced, and its oscillatory properties are more easily stimulated. Second, the frequency of oscillation depends to some degree upon the mechanical damping, and by varying the air gap a variation of the frequency can be obtained, which however is only over a small range. Electrodes must be unaffected by atmospheric conditions and are therefore usually made from stainless steel.

Maintenance of oscillations

61. In order to maintain the crystal in oscillation at its natural frequency, it is necessary to supply power in a manner similar to that employed for the maintenance of any other oscillation, and a thermionic valve (triode) is generally adopted for this purpose. A suitable circuit is shown in fig. 41. The quartz crystal is inserted between the two electrodes of the mounting and the latter are connected to grid and filament of the valve, a grid leak being provided to avoid complete insulation of the grid, and to maintain the latter at the desired potential with reference to the filament. The anode circuit contains an inductance L which may have a condenser C connected in parallel as shown, but for the present this condenser may be disregarded. The existence of the grid-anode inter-electrode capacitance is of importance in the action, which is approximately as follows. Assuming that the filament is heated, on closing the switch (S) there will be an increasing electron current through the valve, and owing to the reactance of the coil L this increase of current

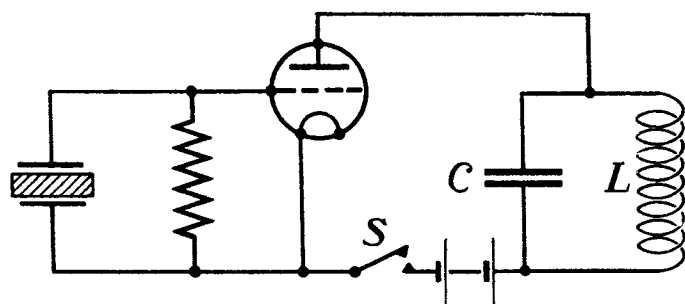
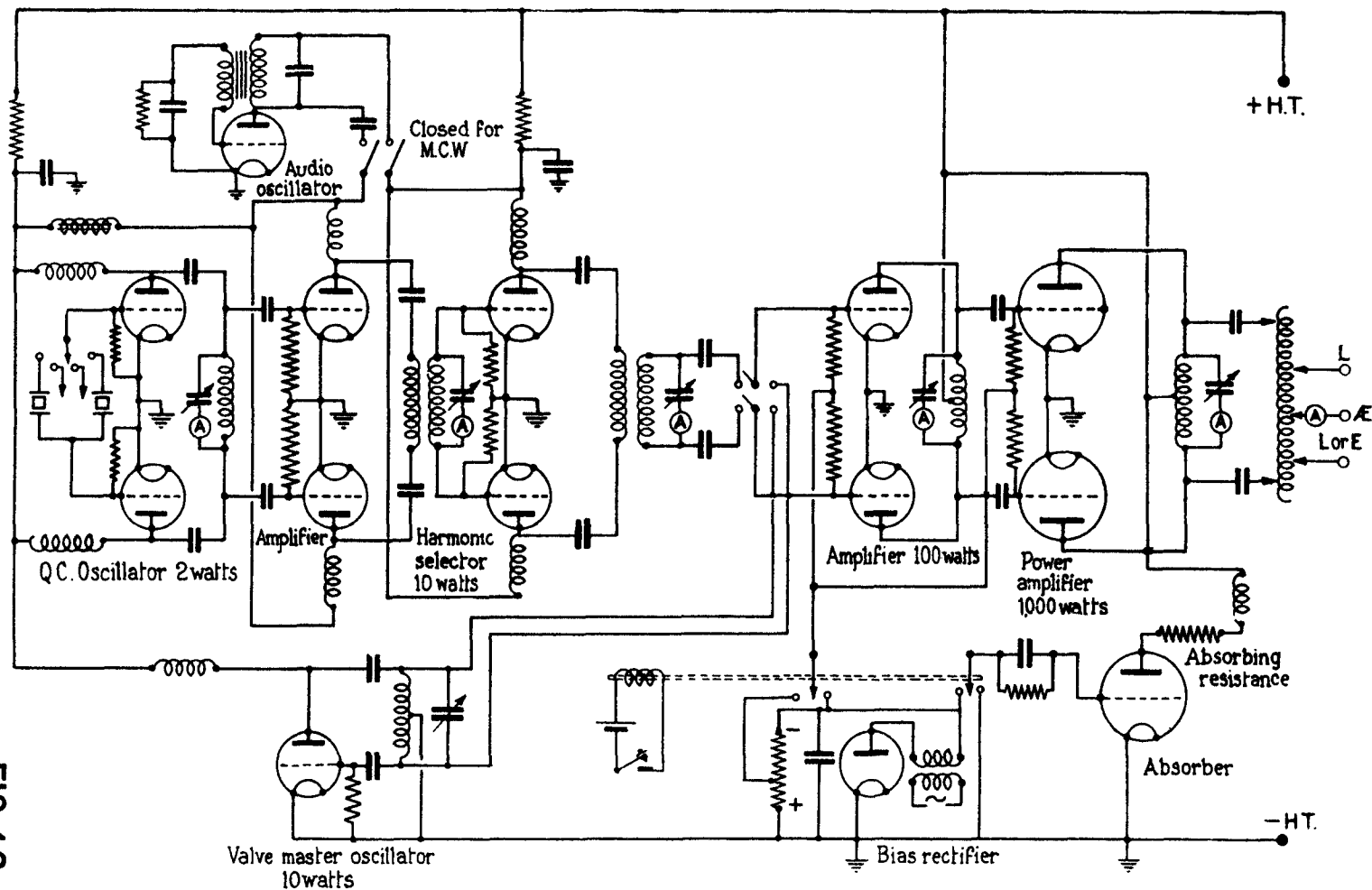


FIG. 41, CHAP. IX.—Circuit in which oscillation may be maintained by quartz crystal.

is accompanied by a fall of anode-filament potential. The grid-anode inter-electrode capacitance and the small condenser formed by the crystal and its electrode are in series between anode and filament, and therefore the fall of anode-filament potential will be accompanied by a change of potential between the electrodes, i.e. a change of P.D. between the faces of the crystal. The latter will therefore either expand or contract, and will then vibrate mechanically with lightly damped sinusoidal motion, the mode of vibration being longitudinal. As it vibrates, sinusoidal variations of P.D. are developed upon its faces, and by means of connections to the electrodes the varying P.D. is applied between grid and filament. The variations of anode current so caused must cause further changes of anode-filament P.D. and consequently, further oscillation of the crystal, the phasing of these resultant changes being such that the crystal is maintained in mechanical vibration at its natural frequency with undamped or increasing amplitude. The correct phasing conditions are fulfilled if the anode circuit reactance is positive, i.e. inductive at the frequency to which the crystal is resonant. It will be observed that the anode current is a pulsating one and may be considered to consist of a steady component and an oscillating component, the latter having the frequency of the crystal oscillation. By means of the condenser C the anode circuit may be brought nearly into resonance with this frequency, with a considerable increase in the circulating current of the circuit. This crystal controlled oscillation may be used to control the frequency of the requisite number of amplifying stages combined if necessary with frequency multiplication, so that the desired aerial output power is achieved.



TYPICAL QUARTZ CONTROLLED TRANSMITTER

Effect of temperature variation

62. X-cut crystals have a negative temperature coefficient, the natural frequency decreasing by about 30 parts in a million per degree Centigrade. They are suitable for use in the type of mounting shown in fig. 40 and generally require a small air gap. In commercial practice they are frequently used to control the frequency of ground transmitters without temperature control, because they oscillate readily and parasitic oscillations are not so troublesome as in Y-cut crystals. Y-cut crystals have a positive temperature coefficient which may be as much as 100 parts in a million per degree Centigrade. If used in a mounting without an air gap, oscillations are more easily maintained than in the X-cut plate. For this reason Y-cut plates are often used in commercial aircraft and marine transmitters operating upon a single frequency. They are more prone to parasitic oscillation than are X-cut plates. If cut from a natural crystal at certain definite angles with respect to the X, Y and Z axes, it is possible to obtain a plate having a temperature coefficient approaching zero. The AT-cut for example, is parallel to the X axis and inclined at an angle of 35° to the Z axis. This angle has to be very accurately determined and maintained during cutting and grinding, and consequently such plates are not in common use.

High frequency transmitters

63. The general principles outlined above are applicable to transmitters up to a frequency of about 60 M.c/s. As the frequency increases beyond about 1 M.c/s, increasing care must be exercised to avoid stray coupling between components and the shunting effect of accidental capacitance. The avoidance of these effects is made more difficult by the necessity for close grouping of the various parts in order to reduce the resistance and inductance of connecting links. Condensers which function perfectly on lower frequencies may become resonant circuits or inductive reactances owing to the presence of their internal conductors. Ohmic losses in conductors may actually be increased by using tube or wire of a gauge larger than a certain optimum and in any case it is desirable that the turns of tuning coils should be spaced by an amount equal to, and preferably double, the diameter of the conductor. Interchangeable coils are a necessity if any considerable frequency range is to be covered, switch or plug variation of inductance being avoided whenever possible. Radio-frequency chokes may give rise to irregularities of operation owing to their distributed capacitance.

As the frequency becomes higher it is a matter of increasing difficulty to achieve the degree of frequency constancy demanded by the ordinary method of C.W. reception, i.e. the heterodyne method described in Chapter X. A frequency variation of 500 cycles per second will cause serious inconvenience to a receiving operator and almost certainly render many repetitions necessary. If the transmitter frequency is only 500 k.c/s. this corresponds to a stability of one part in a thousand, which is easily achieved without frequency control. At 5 M.c/s, however, it is equivalent to one part in ten thousand, which is hardly attainable without frequency control, except possibly in transmitters of very low power. At 50 M.c/s. it corresponds to one part in 100,000, which is a high standard even for a quartz-controlled transmitter with thermostatic temperature control. A partial remedy is to use I.C.W. instead of C.W., the transmitter then emitting a characteristic note. Heterodyne reception is then of course unnecessary but the frequency band occupied by any one line of communication is increased to a corresponding degree, thus giving rise to interference between adjacent channels. It is difficult to grind quartz crystals for frequencies higher than about 6 M.c/s., for the crystal is then so thin that it becomes slightly flexible and cannot be ground accurately flat and parallel. It is therefore usual in crystal-controlled transmitters of high frequency to generate a frequency of the order of 3 M.c/s, the final frequency being attained by a process of frequency multiplication in the amplifier stages. Tourmaline crystals may be used up to about 20 M.c/s, but this material is much more expensive than quartz and its utility is limited by this factor.

64. The circuit diagram of a typical quartz-controlled ground station transmitter is shown in fig. 42. The quartz-controlled master-oscillator is of the push-pull type, and any one of four crystals may be selected by means of a switch, so that four "spot frequencies" can be maintained with a high order of accuracy. A valve-controlled master-oscillator is also provided in case of a

CHAPTER IX.—PARA. 65

breakdown of any crystal, or to generate any frequency other than those provided for by the crystal control. Both the master-oscillators may be thermostatically controlled, although no arrangements to this end are given in the diagram. Following the crystal oscillator is a buffer amplifier having an oscillatory circuit magnetically coupled to its anode circuit; this supplies the excitation to the grid-filament path of the following stage. The oscillatory circuit may be tuned either to the crystal frequency, or to a harmonic, and is referred to as the harmonic selector. As the push-pull layout greatly reduces the even harmonics it is necessary to select and amplify one of the odd ones, the third being generally used. The third and fourth stages are power amplifiers, and must be neutralised, but the neutralising condensers have been omitted for simplicity. The output of the final stage is applied to an auto-transformer which serves to match the anode A.C. resistance of the valves to the actual load, the latter being of course the transmitting aerial. If directly fed from the transmitter, the aerial is connected to the terminal marked (AE) and the earth or counterpoise to the terminal marked (L or E) while if the aerial is fed through transmission lines, these are connected to the terminals (L) and (L or E).

The keying arrangements are shown only in outline, the system used being called "absorption" keying. The object of this is to keep the H.T. voltage as nearly constant as possible, by maintaining an approximately equal load on the generator at all times. The two power amplifying stages derive their negative grid bias from a diode rectifier which is fed from a single-phase A.C. supply, and the magnetic key is arranged to vary the bias. When the key is "up" the grids are biased so negatively that the anode current in both stages is reduced to zero, and when "down" a portion of the bias resistance is short-circuited, the grids become less negative, and normal anode current flows. The absorber valve has a resistance of suitable value in its anode circuit, and the power dissipated in this stage is equal to that dissipated in the power amplifier stages. When the key is "up" the absorber valve grid bias is normal, allowing appreciable anode current to flow, while when "down" the bias is so negative that the anode current is suppressed. In effect therefore, when the morse key is operated, the magnetic key transfers the generator load from the power amplifiers to the absorption valve and vice versa. The actual keying circuits are generally more complex than those shown. The anode voltage of the buffer stage can be varied at a suitable audio frequency by closing the switch marked "close for M.C.W." The transmission is then of the modulated continuous wave type.

Ultra-high-frequency oscillations

65. The method of maintaining oscillations by means of the triode which was discussed in the earlier portion of this chapter, depends upon the production of an oscillatory grid-filament P.D. by coupling the grid circuit to the anode circuit of the triode. This may be termed the "reaction" method of maintaining oscillation. If in a circuit of this type an attempt is made to produce extremely high frequency oscillations, it is found that there is a limit beyond which any given valve will fail to produce the desired effect. This failure is in part caused by the difficulty of designing an oscillatory circuit having a natural frequency higher than 100 M.c/s (corresponding to a wavelength of 3 metres) owing to the inherent inductance of connecting leads to the valve electrodes and to the existence of distributed capacitance. Careful attention to these points may result in an increase in frequency up to about 300 M.c/s ($\lambda = 1$ metre) but it is impossible to maintain oscillations at a frequency appreciably higher than this owing to a further limitation which does not depend upon the circuit constants, namely, the time taken for the emitted electrons to reach the anode of the valve. It may be said that at frequencies sufficiently high to make the mass of the electrons an appreciable factor in the determination of their motion, the reaction method fails to maintain the desired oscillation.

Electronic oscillations are defined as oscillations produced in a thermionic system by virtue of the inertia of the electron. The work of Barkhausen and Kurz is generally regarded as the foundation of this branch of the subject, their discovery of electronic oscillations resulting from certain tests on the degree of vacuum in transmitting valves. In these experiments a positive potential was applied to the grid and a negative potential to the anode, and positive ions produced by grid current would then fall into the anode, so that the magnitude of the anode current could be taken as an indication of the degree of softness of the valve. As a result, it was found that

an anode current was sometimes established in such a direction that it could only be caused by an electron flow to the anode, and not a positive ion current. When this phenomenon was produced, a simple wavemeter situated near the valve indicated the production of oscillations by the valve. Further research shewed that if the grid of a triode is maintained at a positive potential with reference to the filament, while the anode potential is maintained at zero or a slightly negative value, oscillations can be maintained in a circuit connected either between grid and anode, grid and filament, or anode and filament. These oscillations depend upon periodic motion of electrons in the inter-electrode space, and are only maintained when the natural frequency of the circuit has a certain relationship with the frequency of vibration of the electrons. One explanation advanced by these physicists to explain the results is as follows. Electrons emitted by the filament receive acceleration owing to the positive grid potential, and some of them pass through the grid, after which their acceleration becomes negative. A number of them therefore have their direction reversed, travelling back towards the grid, passing through the latter to the vicinity of the filament. Here they are repelled by the presence of other emitted electrons and attracted by the positive grid, their direction of motion again undergoing reversal. Hence a number of electrons are constantly maintained in periodic or oscillatory motion in the inter-electrode space, and this constitutes an oscillatory current of very high frequency. With certain assumptions, the wavelength can be calculated from the formula

$$\lambda = \frac{1,000}{\sqrt{V_g}} \frac{d_a V_g - d_g V_a}{V_g - V_a}$$

in which

d_a = twice the distance between anode and filament.

d_g = twice the distance between grid and filament.

V_a = anode-filament P.D.

V_g = grid-filament P.D.

One of the assumptions made in deriving this expression is that the electrodes are flat parallel planes, and the error in using this formula to calculate the wavelength obtainable with triodes of ordinary design may be as much as 25 per cent.

66. The above theory fails to account for the establishment of an electron current in the anode circuit of the valve. It has been pointed out, however, that if an alternating component of P.D. exists between grid and anode, work is done on the moving electrons during the half-cycle in which they receive acceleration toward the anode, and work is done by the electrons during the half-cycle in which they move toward the grid. Electrons reach the anode during the former half-cycle, if the steady negative potential is less than the peak value of anode grid P.D. on the positive half-cycle. The work done by the electrons determines the amount of damping which can be made good in the oscillatory circuit, resembling in this respect the overcoming of

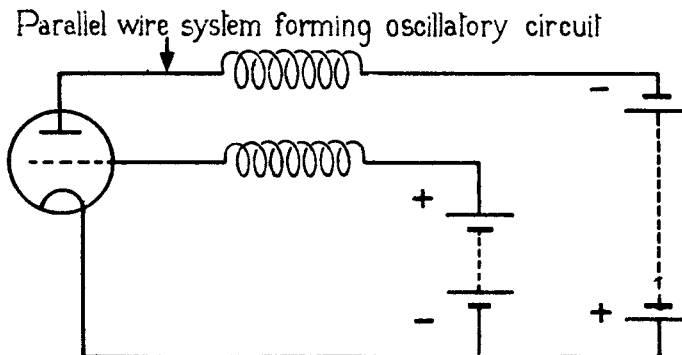


FIG. 43, CHAP. IX.—Barkhausen-Kurz oscillator.

CHAPTER IX.—PARA. 67

damping by reaction in the ordinary triode oscillator. From this point of view, the anode current represents electrons which are wasted so far as the production of oscillations is concerned, i.e. it represents a form of damping loss. Another mechanism of maintenance is possible when the anode diameter is considerably greater than the grid diameter and the anode is at filament potential. The grid-anode space may then become saturated while the grid filament space is unsaturated, and it has been suggested that successive regions of high and low electron density may be set up in the grid-anode space, these regions moving to and fro periodically in a manner resembling the compression and rarefactions constituting a sound wave. A circuit connected between grid and anode, the natural period of which is twice the time taken by an electron to pass from grid to anode, would then be maintained in oscillation. A circuit suitable for the production of B-K oscillations is shown in fig. 43, although all triodes are not adapted for use in this manner. Best results are obtained with an optimum ratio of anode diameter to grid diameter, which is about 3 : 1. A close spiral of fine wire appears to form the design of grid most conducive to the maintenance of oscillations, while the anode should be of cylindrical form. Attempts to produce the desired results with the commercial designs of valve having an anode of flattened box form with a V or N filament have so far been unsuccessful, and best results are obtained with valves in which the filament is short and straight, with the grid and anode carefully mounted concentrically with the filament. Only by such precautions can any appreciable power output be obtained. The grid becomes the electrode which is called upon to dissipate most heat, and grid failure is usually the factor which determines the life of the valve.

It has been stated that the anode electron current represents a damping loss, and it is desirable to reduce this as far as possible. This is the chief reason for applying a negative potential to the anode, although the equation given above shews that an increase in negative anode potential leads to a slight decrease of wavelength or increase of frequency. It may now be stated that this equation is based on the assumption that the period of the oscillation is approximately four times the "time of flight" of an electron, the latter term being now in common use to denote the time of transit of an electron from cathode to anode.

The magnetron

67. The generic term magnetron is applied to any type of thermionic valve in which the anode current is controlled by the presence of a magnetic field, instead of, or sometimes in addition to, the control exercised by the electric potential of the grid. The practical form of magnetron is usually a diode and only this form need be mentioned. A typical example is shown diagrammatically in fig. 44 in which the cathode is a directly heated filament, and is surrounded by a cylindrical anode, which is split along its length to reduce eddy current losses. The electrodes are mounted in a glass envelope in the usual way, but in the anode-cathode space a magnetic field of about 20 to 50 oersteds is maintained by a solenoidal winding which surrounds the bulb. The figure shews that the filament lies in the same direction as the magnetic field. The electrons

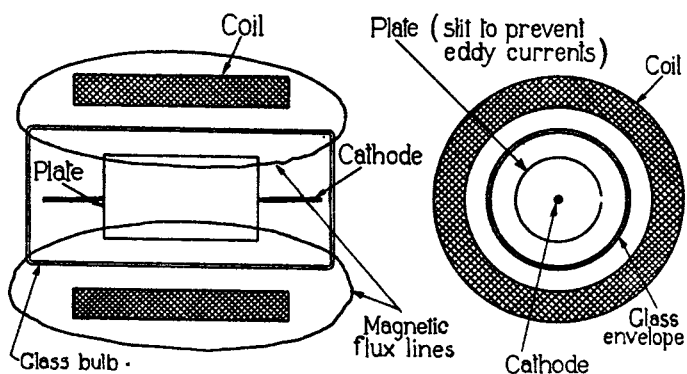


FIG. 44, CHAP. IX.—Diagrammatic representation of magnetron

which are emitted by the filament constitute an electric current, and on several previous occasions reference has been made to the force exerted by a magnetic field upon an electric current—or upon a moving electric charge, which is the same thing. An electron which is emitted from a filament into a magnetic field will experience this force, and consequently receive acceleration, the latter being proportional to its velocity and at right angles to its direction of motion. If the magnetic field is uniform this results in the electron taking a path which is the arc of a circle. This effect is present in all valves having a co-cylindrical arrangement of directly heated filament and anode, owing to the magnetic field set up by the filament heating current, but it only becomes of importance in high-power valves carrying filament currents of the order of 50 amperes, and

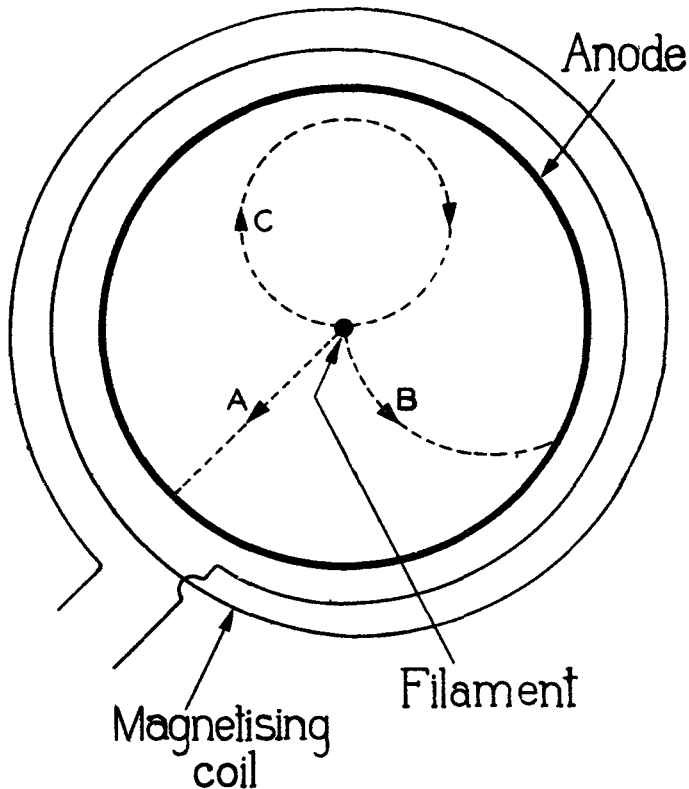


FIG. 45, CHAP. IX.—Paths of electrons with variation of magnetic field strength.

the additional magnetising coil is necessary to produce magnetron phenomena in low-power valves. The effect of the magnetic field upon the anode current is shewn in fig. 45. If the magnetic field strength is negligible an electron leaving the filament will travel to the anode by the shortest path, as shewn by the dotted line A. An increase in the strength of the field will cause the electron to be deflected, travelling a path somewhat as shewn at B, while if the field strength exceeds a certain value the electron will travel in a closed path as shewn at C, in which case the anode current will fall to zero. The magnetic field strength necessary to cause the electrons to fail to reach the anode is given by the equation

$$H = \frac{6.72\sqrt{V}}{r}$$

CHAPTER IX.—PARA. 68

where V is the anode-filament P.D. and r the radius of the anode. This "cut-off" is very critical, and a considerable variation of anode current is obtainable for slight variations of magnetic field strength. The action of a magnetron is therefore analogous with that of a triode, except that the anode current is controlled by variation of the strength of the magnetic field instead of by variation of the grid-filament potential.

68. The type of magnetron most used for the production of very high frequency oscillations is the split-anode type shewn in fig. 46. The anode consists of two half-cylinders of nickel which are insulated from each other and have separate leading-in conductors. A tuned circuit is con-

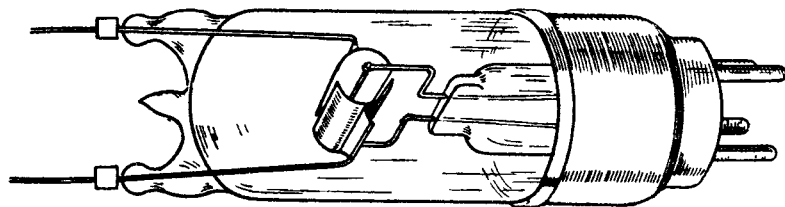


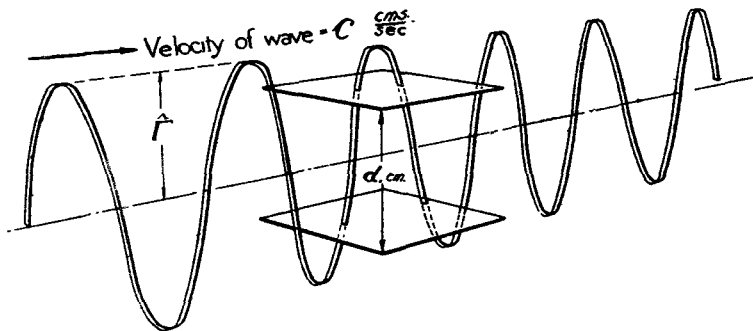
FIG. 46, CHAP. IX.—Split-anode magnetron.

nected between the two halves of the anode, the H.T. supply being arranged by means of a centre tap on the inductance, the magnetic field being produced by a solenoidal winding carrying direct current, and the field strength being adjusted to a value at which the anode current is reduced to nearly but not quite zero. The mechanism by which oscillations are produced is not entirely understood, but the magnetron is certainly capable of maintaining oscillations of the order of 5,000 M.c/s, corresponding to a wavelength of about six centimetres, and the power obtainable is much greater than with triodes producing the Barkhausen-Kurz type of oscillation. It is therefore probable that the magnetron will undergo considerable development to meet the demand for additional communication channels in the ultra-high frequency band.

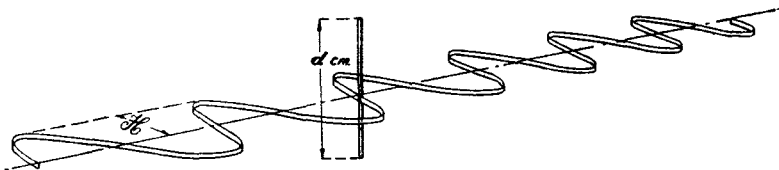
CHAPTER X.—RADIO TELEGRAPHIC RECEPTION

E.M.F. AND CURRENT IN A RECEIVING AERIAL

1. In the preceding chapters we have discussed the production of electromagnetic radiation of frequencies such as are actually employed for radio-communication, without reference to the means employed to detect the presence of such radiation at the receiving station. The principles involved in the latter process will be the subject of the present chapter. We have already seen (Chapter VII) that the radiation from a transmitting aerial consists of a travelling electrical field, the strength of which varies sinusoidally both in space and time, or alternatively of a magnetic field having similar characteristics. These fields are in phase with each other, and if the wave is normally polarised, the direction of the electric field is vertical and that of the magnetic field horizontal, the plane of polarisation being perpendicular to the direction of propagation of the wave. This electromagnetic field travels with a velocity of 3×10^{10} centimetres per second, which is generally denoted by the symbol c . Consider the effect of the electromagnetic wave upon any electric circuit in its path. If at any point in the field two metal plates are suspended one above the other, d centimetres apart (fig. 1a), the passage of the electromagnetic disturbance through the volume of ether enclosed by them is equivalent to the existence of a P.D. between the plates, and we may find the magnitude of this P.D. quite simply. If the peak electric field strength is E electrostatic units, the peak P.D. is equal to $E d$ E.S. units or $300 E d$ volts, because 1 E.S. unit of P.D. is 300 volts. Instead of suspending this condenser in mid-air, consider a single vertical wire d centimetres in length, as in fig. 1b. A magnetic field of strength H travelling with a velocity of c centimetres per second in the region in which the wire is situated, will set up an induced E.M.F. in this wire



(a) P.D. set up between condenser plates



(b) E.M.F. induced in vertical wire

FIG. 1, CHAP. X.—Induced voltage due to electromagnetic field.

CHAPTER X.—FARA. 2

in accordance with Faraday's law, and again its magnitude can be easily calculated, from the relation $\mathcal{E} = \frac{c\mathcal{H}d}{10^8} = 300\mathcal{H}d$ volts. As previously shewn (Chapter VII) the magnetic field strength $\mathcal{H}$ and the electric field strength $\hat{F}$ of the wave are equal numerically although expressed in different units, and therefore, whether we chose to consider the wave as an electrical or as a magnetic phenomenon, the magnitude of the E.M.F. set up in a given region is the same. It must be carefully noted that the voltage at any point in the field is not the sum of the two voltages derived above, for these are merely different aspects of the same thing. The electric field is identical with the moving magnetic flux, while the magnetic field is identical with the moving electric flux, and the peak P.D. between two points in space, their vertical distance apart being d centimetres, is either $300 \hat{F}d$ or $300 \mathcal{H}d$ volts. At the distances over which radio communication is generally used, the electric field strength is only of the order of from 10^{-7} to 10^{-9} E.S. Units. It is therefore more conveniently expressed in millivolts (or microvolts) per metre: Since $1 \text{ E.S.U.} = 300 \text{ volts per centimetre}$, $1 \text{ millivolt per metre} = \frac{1}{3 \times 10^7} \text{ E.S.U.}$

Example.—At Cranwell, the peak value of the electric field strength due to the London Regional transmitter of the B.B.C. was found to be 10^{-8} E.S.U. Express this in millivolts per metre and find the peak E.M.F. induced in a vertical aerial 10 metres in height.

$$\begin{aligned} 1 \text{ E.S.U.} &= 300 \text{ volts per centimetre} \\ \therefore 10^{-8} \text{ E.S.U.} &= 3 \times 10^{-6} \text{ volts per centimetre} \\ &= 3 \times 10^{-4} \text{ volts per metre} \\ &= \cdot 3 \text{ millivolts per metre.} \end{aligned}$$

The total E.M.F. acting in the aerial circuit will therefore be $\cdot 3 \times 10 = 3$ millivolts.

2. As the vertical wire must possess distributed capacitance, inductance and resistance, the induced E.M.F. is able to establish a current in the wire, and the latter is then said to act as a receiving aerial. The current may be considered to be a conduction current in the wire, and a corresponding displacement current in the surrounding medium, the circuit consisting of the wire itself and the capacitance existing between different parts of the conductor. If the induced voltage has an R.M.S. value of E volts, and the total impedance of the aerial circuit is Z ohms, the R.M.S. value of the current is given by the usual equation,

$$I = \frac{E}{Z}$$

and will attain its greatest value if the receiving aerial circuit is in resonance with the frequency of the distant radiant circuit, i.e., the transmitter. The receiving aerial will offer no reactance at this frequency and the value of the current will be $\frac{E}{R}$. In order to obtain maximum current, therefore, the inductance and capacitance in the receiving aerial circuit must be adjusted to such values that

$$\omega L = \frac{1}{\omega C}$$

where L and C are the effective inductance and capacitance of the whole circuit. The principles involved in tuning receiving aerial to a given frequency may be simplified by considering only the earthed aerial, and assuming also that its effective capacitance C and inductance L are constant for all frequencies. Thus a certain small receiving aerial is found to have an effective inductance of 60 microhenries, an effective capacitance of 170 micro-microfarads, and an effective

resistance of 50 ohms. With neither inductance nor capacitance added, therefore, its resonant frequency is

$$\begin{aligned}
 f_r &= \frac{1}{2\pi\sqrt{LC}} \\
 &= \frac{1}{2\pi\sqrt{\frac{60}{10^6} \times \frac{170}{10^{12}}}} \\
 &= 1.57 \times 10^6 \text{ cycles per second (1,570 k.c/s.)}
 \end{aligned}$$

Aerial tuning

3. (i) Now suppose it is required to adjust this aerial to receive a signal on a frequency of 8×10^5 cycles per second. As the signal frequency is lower than the resonant frequency of the aerial, the desired result is obtained by adding inductive reactance in series between the aerial and earth. To find the reactance required, we observe that the reactance of the aerial alone is

$$\begin{aligned}
 X_a &= \omega L - \frac{1}{\omega C} \\
 \text{and } \omega &= 2\pi \times .8 \times 10^6 = 5.025 \times 10^6 \\
 X_a &= (5.025 \times 10^6 \times 60 \times 10^{-6}) - \left(\frac{10^{12}}{5.025 \times 10^6 \times 170} \right) \\
 &= 301.5 - 1168 \\
 &= -866.5 \text{ ohms.}
 \end{aligned}$$

The negative sign confirms that the reactance of the aerial circuit at the particular frequency is capacitive, and an inductive reactance of + 866.5 ohms must be added to bring the aerial circuit into resonance with the distant transmitter. This corresponds to an inductance of 172 microhenries, although (as shown in Chapter V) the added reactance may be constituted by an inductance of smaller value than this, in parallel with a condenser of a certain value. The cardinal point to be remembered is that if the distant transmitter has a frequency higher than the natural frequency of the aerial, capacitive reactance must be interposed between the

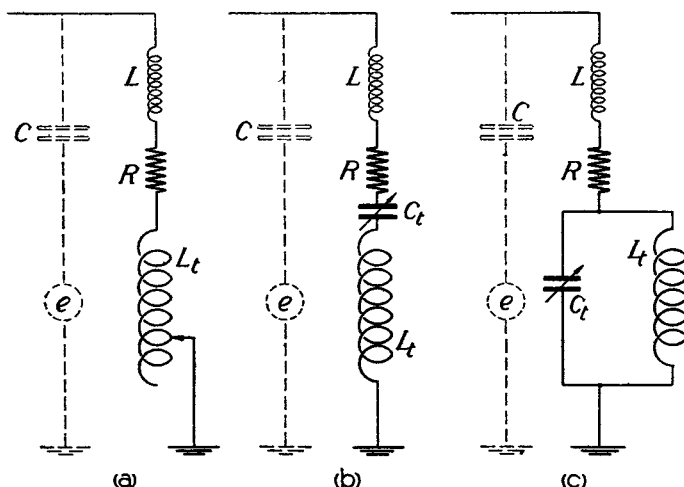


FIG. 27. CHAP. X.—Tuning devices for receiving aerial.

CHAPTER X.—PARA. 3

aerial and earth, while if the frequency of the transmitter is lower than the natural frequency of the aerial, inductive reactance must be inserted. The necessary tuning reactance may be obtained by a combination of inductance and capacitance, either or both being variable in order that the receiver may be adjusted to cover a given frequency range, but no matter how complex the aerial circuit may appear, the aerial circuit as a whole always operates as an acceptor circuit for the frequency to which it is tuned. Three possible arrangements of the receiving aerial circuit are shown in fig. 2, in which L , C , R are the constants of the aerial, ϵ the effective E.M.F. induced in the aerial by the incoming signal, and L_t , C_t the tuning inductance and capacitance.

(ii) The arrangement shown in fig. 2a has already been illustrated by a numerical example; it should be noted that if it is desired to receive frequencies above the resonant frequency of the aerial, the total tuning reactance must be capacitive, and it will be necessary to insert a series condenser, resulting in the circuit shewn in fig. 2b.

Example.—In fig. 2b if the aerial inductance is $60 \mu H$, its capacitance $170 \mu\mu F$, and the added inductance $1080 \mu H$, find the frequency range which will be covered if C_t is variable between 30 and $300 \mu\mu F$.

The total effective capacitance C_e is that of C and C_t in series, or $C_e = \frac{C C_t}{C + C_t}$.

When $C_t = 30 \mu\mu F$,

$$C_e = \frac{30 \times 170}{30 + 170} = 25.5 \mu\mu F.$$

When

$$C_t = 300 \mu\mu F,$$

$$C_e = \frac{300 \times 170}{300 + 170} = 108.5 \mu\mu F.$$

The lowest frequency will be

$$f_l = \frac{10^9}{2\pi\sqrt{(1080 + 60) \times 108.5}}$$

$$= 451,000 \text{ cycles per second (451 k.c/s.)}$$

and the highest frequency

$$f_h = \frac{10^9}{2\pi\sqrt{(1080 + 60) \times 25.5}}$$

$$= 930,000 \text{ cycles per second (930 k.c/s.)}$$

(iii) In the arrangement shown in fig. 2c it is easily demonstrated by methods explained in Chapter V that if L_t is very much larger than L the resonant frequency is approximately

$$f_r = \frac{1}{2\pi\sqrt{L_t(C + C_t)}}$$

i.e., for tuning purposes we must regard the tuning condenser to be in parallel with the aerial capacitance. It must particularly be noted that no "rejector" action is involved in the parallel combination of L_t and C_t since it is not in resonance with the signal frequency.

Example.—If the aerial inductance is $60 \mu H$, the aerial capacitance $170 \mu\mu F$, the added inductance $1080 \mu H$ and the added capacitance is variable from 30 to $300 \mu\mu F$, find the tuning range.

The highest frequency will be

$$f_h \doteq \frac{10^9}{2\pi\sqrt{1080(170 + 30)}}$$

$$= 334,200 \text{ cycles per second (334 k.c/s.)}$$

and the lowest frequency

$$f_1 \doteq \frac{10^9}{2\pi\sqrt{1080(170+300)}} \\ = 224,000 \text{ cycles per second (224 k.c/s.).}$$

(iv) In obtaining the above results, it was assumed that the aerial capacitance is constant at all frequencies, but this is not strictly true; nevertheless, the approximations are sufficiently close to enable one to predict the probable tuning range of a given circuit when used with a particular aerial. Other arrangements of the aerial tuning system are discussed in later paragraphs, and immediate attention must be devoted to the means whereby the reception of a signal is made perceptible to the human senses. In order to appreciate this problem it is essential to bear in mind that the strength of the field radiated by a transmitter is rapidly attenuated as the distance from the transmitter increases and it is often necessary to detect the presence of a signal having a field strength of the order of one microvolt per metre. If the height of the receiving aerial is 10 metres and its resistance, including that of the tuning devices and earth connection, 50 ohms, a field of this intensity would establish an aerial current of only $\frac{10 \times 1 \text{ microvolt}}{50 \text{ ohms}} = \cdot 2 \text{ microampere}$. It is quite impossible to detect an alternating current

of this magnitude by any direct-reading instrument such as a hot-wire ammeter and it is necessary to employ some form of detector which possesses the required degree of sensitivity.

RECTIFICATION

Necessity for rectification

4. The telephone receiver described in Chapter II can be designed in such a manner that it is sensitive to extremely small variations of current, i.e. of the order of one-tenth of a microampere. At first sight, therefore, it might be inferred that the presence of current in the receiving aerial could be detected by connecting a telephone receiver (or a pair of such receivers) in series between the aerial proper and earth. This simple solution is not effective for several reasons which must be fully appreciated, and therefore merit a somewhat detailed discussion. It is convenient to refer to the physiological processes by which the sensation of sound is produced in the human brain as "hearing," while the sound may be defined as that which is heard. The process of hearing is extremely complicated and, like all other phenomena connected with the nervous system, cannot be said to be fully understood. It is however known that when the drum of the ear is set into vibration by sound, the presence of the latter is signalled to the brain. The sound consists of mechanical vibration of the medium with which the ear-drum is in contact, which is generally air at atmospheric pressure, and the sensations produced by the sound depend upon three characteristic properties of the vibration. These are first the frequency, second the intensity and third the wave form, but when dealing with C.W. or I.C.W. reception we are chiefly concerned with the first of these characteristics, the frequency. The ear and brain perceive as sound only such vibrations as possess frequencies lying between a lower and an upper limit; these limits cannot be sharply defined, for they vary in different individuals, and even in the same individual according to the circumstances existing at the time. Direct detection of a radio signal by the telephone receiver, then, is not possible for the following reasons:—

(i) In normal individuals the extreme lower limit is about 16 cycles per second, and the upper about 20,000 cycles per second. The upper limit is therefore lower than the lowest frequency used for radio communication, and consequently, if a telephone receiver were connected in the manner suggested above, the vibration of the telephone diaphragm at the frequency of the signal would not convey to the brain any sensation of sound.

(ii) In a telephone receiver of ordinary design, the passage of a radio-frequency current would not set up an appreciable vibration of the diaphragm, which owing to its stiffness and inertia possesses a natural frequency of the order of only 1,000 cycles per second. The resonance is not very sharp in the normal designs, and a fairly even response can be attained over a frequency

CHAPTER X.—PARA. 5

range of say 200 to 2,000 cycles per second, but the response of the diaphragm falls off rapidly as the latter frequency is exceeded, and above about 5,000 cycles per second practically no response can be obtained. A typical response characteristic for a commercial instrument is shewn in fig. 3.

(iii) The windings and connecting leads of the telephone receiver possess considerable inductance and distributed capacitance. The inductance of a telephone winding of ordinary design may be of the order of 1 henry, which is very much greater than is required in the aerial circuit for tuning purposes. This inductive winding is however shunted by the distributed capacitance which may be of the order of $100\mu\mu F$ and consequently will offer very much less opposition to an alternating current of radio frequency, e.g. if $\omega = 5 \times 10^6$ the capacitive reactance of the shunt path is only 2×10^4 ohms, while an inductance of 1 henry offers a reactance of 5×10^6 ohms. Hence only a small portion of the current will flow through the windings, and very little response would be obtained even if the mechanical characteristics of the diaphragm could be modified to allow it to vibrate at higher frequencies.

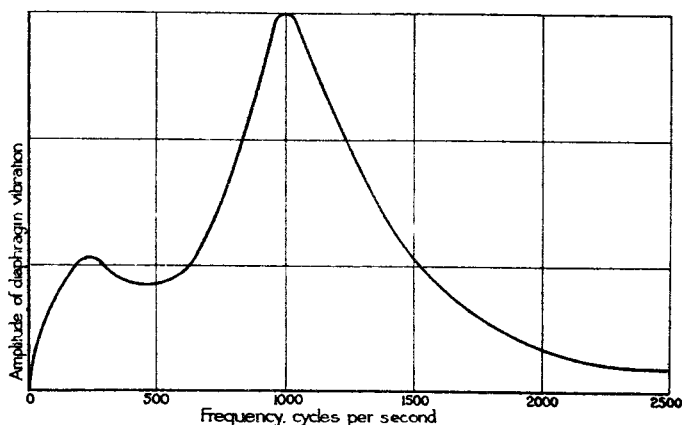


FIG. 3, CHAP. X.—Frequency response of telephone receiver.

Rectification by ideal diode

5 (i) It may appear that the impossibility of direct employment of the telephone receiver for radio reception has been given more consideration than is merited by its importance, but experience shows that operators frequently fail to appreciate the necessity for the inclusion of other apparatus in the receiver. The telephone receiver can be, and is, almost universally used as a receiving device in radio-telegraphy, in conjunction with other devices which cause the current through the telephone to vary in amplitude at a frequency which is within its responsive capability. The necessary steps are first, rectification of the radio-frequency currents and second, variation of their amplitude at a rate which is within the limits of efficient response of the telephone, e.g. from 200 to 2,000 times per second. A rectifier has already been defined as a conductor which does not obey Ohm's law, or which has a non-linear current/voltage characteristic, and the rectification of an alternating voltage in order to provide direct current H.T. supply for the C.W. transmitter has been described in the previous chapter. Now suppose that owing to the operation of a distant C.W. transmitter a radio-frequency current is established in a receiving aerial circuit which is tuned to the transmitter frequency by means of an added inductance; as shewn in Chapter V, the P.D. across the ends of this inductance may be much greater than the E.M.F. induced in the aerial. A rectifier and a telephone receiver, in series, may be connected across the ends of the inductance, the circuit being then as shown in fig. 4, in which the rectifier is a two-electrode valve or diode. A

reservoir condenser C_r is connected in parallel with the telephones. This serves not only as a reservoir but also to reduce the radio-frequency impedance of the path between the points A and B to a negligible value.

(ii) The action of this circuit may be studied with the aid of the characteristic curve of the rectifier. The general shape of the diode characteristic has been discussed in Chapter VIII, but it will simplify the present explanation if we assume the diode to possess the ideal characteristic shown in fig. 5, i.e. a linear relation between anode voltage (V) and anode current (I) for all positive values of V , and zero current for all negative values of V . When no oscillatory P.D. exists between the ends of the inductance the anode-filament P.D. of the diode is zero, and no anode current will flow. During the reception of an electro-magnetic wave an oscillatory aerial current will be established, causing a difference of potential between the ends of the tuning inductance. This P.D. is applied to the rectifying circuit, that is, to all intents and purposes, to the anode-filament path of the diode, for as already pointed out, the presence of the reservoir condenser ensures that no appreciable radio-frequency P.D. will exist between the points A and B.

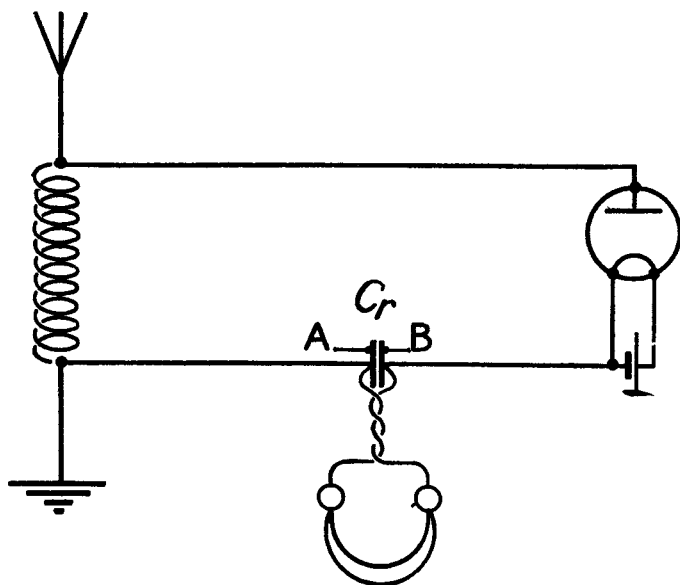


FIG. 4, CHAP X.—Simple receiver with diode rectifier.

During those half-cycles of oscillatory anode-filament P.D. in which the anode is positive with respect to the filament, an anode current will be established, which will flow through the tuning inductance and charge the reservoir condenser. Electrons reaching the plate A set up a displacement current in the dielectric and repulsion of electrons from the plate B, which therefore becomes positively charged. During the half-cycles in which the anode is negative to the filament, the anode filament path of the diode is non-conductive and no such anode current can exist. The electron current flowing into the reservoir condenser at the point A thus consists of a series of unidirectional impulses as shown in the diagram.

6. (i) So far the presence of the telephones has been ignored, for as previously stated no appreciable portion of the radio-frequency currents will pass through them. The windings however constitute a path by which the reservoir condenser is free to discharge, and therefore the charge introduced into it by the impulses of electron current will leak away at the average rate at which it is received, in the form of a unidirectional current through the windings. With

CHAPTER X.—PARA. 6

the ideal form of rectifier characteristic postulated, if the peak voltage applied to the rectifier is $\mathcal{V}$, and the anode A.C. resistance of the valve is r_a , the peak value of the anode current is $\frac{\mathcal{V}}{r_a}$, and the average value of anode current over a complete cycle is $\frac{\mathcal{V}}{\pi r_a}$. This is the value of the steady current which will be established through the windings of the telephone receiver, during the first cycle of the oscillatory voltage. If the impedance of the telephone receiver is negligible compared to the anode A.C. resistance of the valve, each successive positive half-cycle will give rise to a current of this value, so that during the whole of the period in which the oscillatory P.D. is applied, a steady current of $\frac{\mathcal{V}}{\pi r_a}$ amperes will flow through the telephone windings.

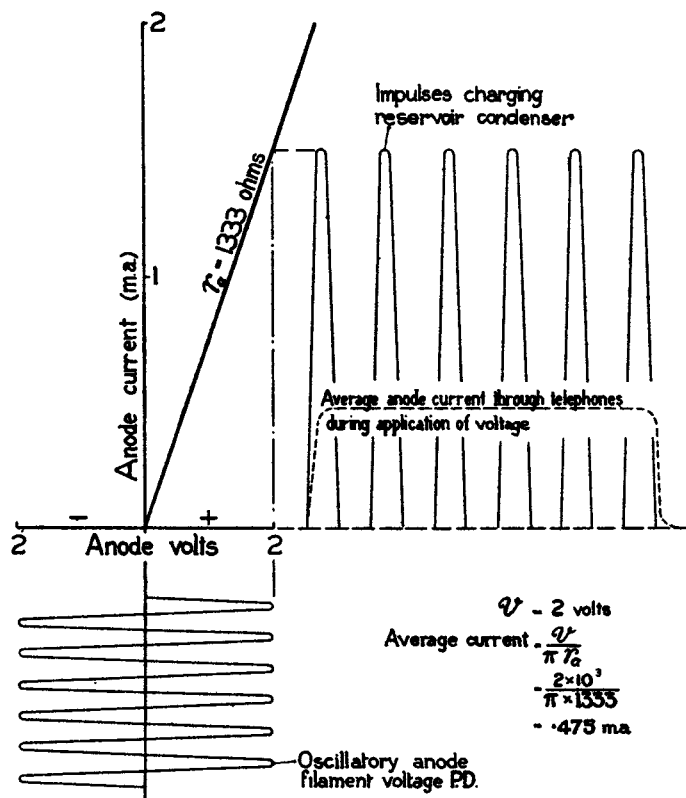


FIG. 5, CHAP. X.—Rectification of oscillatory voltage of constant amplitude by ideal rectifier.

(ii) The effect of this current on the telephone magnets will depend upon the direction of the current, for if in one direction round the winding it will increase the pull on the diaphragm, in the contrary direction it will weaken the magnets slightly and tend to release the diaphragm. In either event, no vibration of the diaphragm ensues, and little if any indication of the presence of the aerial current will be given although if the transmitter is very close to the receiver, a single click may be heard when the transmitting key is pressed, and another when it is released. If however, a high resistance microammeter were inserted in place of the telephones the pointer would be deflected, rising to some mean value, e.g. under the operating conditions of fig. 5, to 475 microamperes.

Approximate rectified current with resistance load

7. The value $\frac{\mathcal{V}}{\pi r_a}$ which has been obtained above assumes that the impedance of the circuit (external to the diode itself), is zero. Since this cannot be true in practice, owing to the presence of the telephone receivers and reservoir condenser, the rectifier current will be less than this, because the P.D. across the reservoir condenser causes the anode to become more and more negative with respect to the filament, during the time that the radio-frequency voltage is applied. The value $\frac{\mathcal{V}}{\pi r_a}$ for the rectified current must therefore be regarded as an upper limit which cannot be exceeded.

An exact derivation of the correct value involves somewhat heavy mathematics and is purely of academic interest, since no rectifier possesses the ideal characteristics postulated. Since, however, it is found experimentally that on the application of a radio-frequency voltage of constant amplitude, the counter-E.M.F. of the reservoir condenser reaches a mean steady value in the course of the first few cycles, continuing then to receive during each successive positive half-cycle a charge just sufficient to maintain a practically steady current through the load, we may obtain an approximation to the load current and the counter-E.M.F. of the condenser as follows. As before, let $\mathcal{V}$ be the peak value of the applied voltage, C_r the capacitance of the reservoir condenser, r_a the anode A.C. resistance of the valve, R the load resistance and e the counter-E.M.F. of the reservoir condenser. Then the charging current during any positive half-cycle will be

$$I = \frac{\mathcal{V} - e}{\pi r_a}$$

flowing for nearly $\frac{1}{2f}$ second. The charge received by the condenser will be

$$Q = \frac{I}{2f} = \frac{\mathcal{V} - e}{2\pi f r_a} = \frac{\mathcal{V} - e}{\omega r_a}$$

and the counter-E.M.F. developed, if the load resistance were infinitely high, would be

$$e = \frac{Q}{C_r} = \frac{\mathcal{V} - e}{\omega C_r r_a}$$

Owing to the presence of the load resistance, the condenser charge leaks away during the negative half-cycles, the average current through the resistance being $\frac{e}{R}$ amperes. During one half-cycle, the charge lost will be $\frac{e}{2fR}$ coulombs, leading to a fall of P.D. equal to $\frac{e}{2fC_r R}$ or $\frac{\pi e}{\omega C_r R}$ volts. Hence

$$\begin{aligned} e &= \frac{\mathcal{V}}{\omega C_r r_a} - \frac{e}{\omega C_r r_a} - \frac{\pi e}{\omega C_r R} \\ e \left(1 + \frac{1}{\omega C_r r_a} + \frac{\pi}{\omega C_r R} \right) &= \frac{\mathcal{V}}{\omega C_r r_a} \\ \text{or } e &= \frac{R \mathcal{V}}{\pi r_a + R + \omega C_r R r_a} \end{aligned}$$

The current I_r through the load resistance is $\frac{e}{R}$, and

$$I_r = \frac{\mathcal{V}}{\pi r_a + R + \omega C_r R r_a}$$

CHAPTER X.—PARA. 8

Actually, the load current will be rather less than this, because the condenser may charge for appreciably less than $\frac{1}{2f}$ second. The average input resistance of the rectifier is also of interest. To the degree of accuracy of the above derivation, if R_0 is the input resistance, since the average value of the applied voltage during one half-cycle is $\frac{2\mathcal{V}^*}{\pi}$, it follows that

$$\frac{2\mathcal{V}^*}{\pi R_0} = \frac{\mathcal{V}^* - e}{\pi r_a}$$

$$\therefore R_0 = \frac{2r_a}{1 - \frac{e}{\mathcal{V}^*}}$$

The average input resistance is of importance in a particular service application of the diode (see paragraph 49).

Detection of I.C.W. signals

8. To produce an audible response in the telephone receiver, it is necessary to interrupt or vary the current at a rate to which the diaphragm will respond. This variation or interruption may be performed either at the transmitter or at the receiver, but if performed at the transmitter

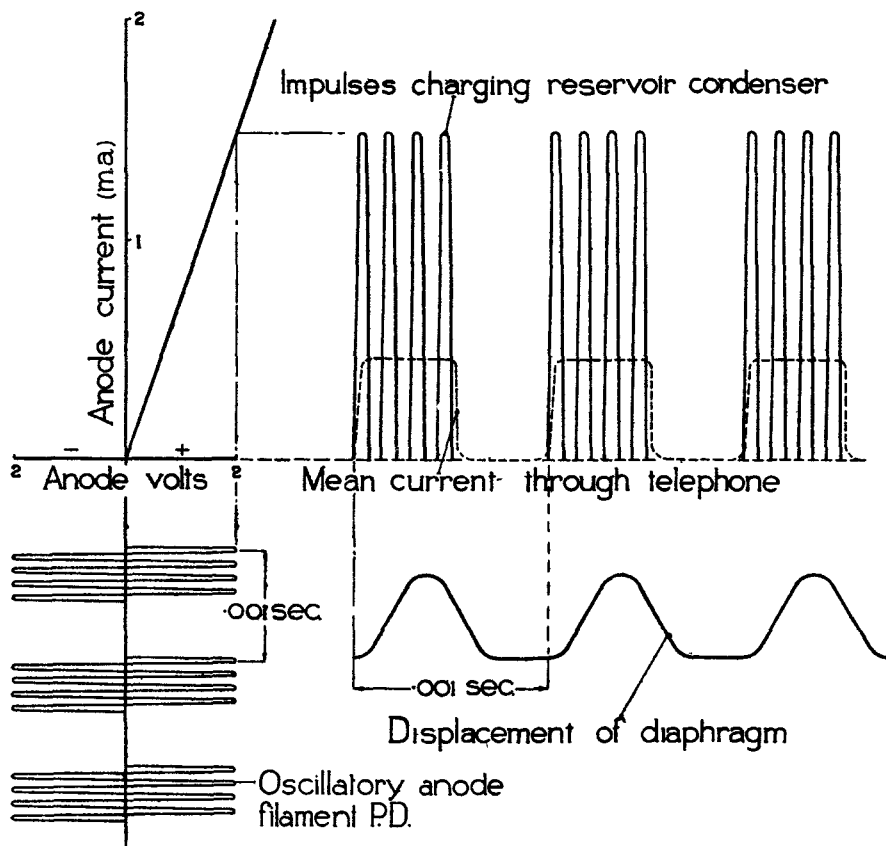


FIG. 6, CHAP. X.—Rectification of I.C.W. signal by ideal rectifier.

the emission is no longer a continuous wave, but some form of modulated wave, i.e., I.C.W., M.C.W., or radio-telephony depending upon the nature of the variation. The receiver action in the case of I.C.W. transmission is shown in fig. 6, which is similar to fig. 5 except that the voltage applied to the rectifier has a different wave form. The transmission may consist of radio-frequency waves which are interrupted 1,000 times per second; if the radio frequency is 1 megacycle per second, one group of 500 complete cycles will occupy $\frac{1}{2000}$ th of a second, while an interval of similar duration will occur between successive groups. As a morse "dot" lasts for about one-twentieth of a second (at 20 words per minute) the number of groups comprising a single "dot" is $\frac{1}{20} \div \frac{1}{2000} = 50$. It will be appreciated therefore that neither the number of waves per group, nor the duration of individual waves and complete groups respectively, can be drawn to scale. During every group of waves, a negative charge will be established upon the plate A of the reservoir condenser, and a corresponding positive charge on the opposing plate B, the condenser however discharging through the telephones at the average rate at which it is charging. The telephone current therefore rises to a value of $\frac{\mathcal{V}}{\pi r_a + R}$ (approximately) during each group of waves and falls to zero when the group ceases; as each group of waves lasts only about $\frac{1}{2000}$ th of a second and is followed by an interval of about the same duration before the commencement of the next group, the telephone current will rise and fall at the same rate, and the telephone diaphragm will be set into vibration at a fundamental frequency equal to the number of groups per second, producing a musical note which is easily distinguished from other noises which may be present (*see* paragraphs 5 *et seq.*). Actually the vibration will contain higher harmonics, but this is of no importance in telegraphic reception.

Square law rectification

9. Having established the general nature of the process of rectification it must now be pointed out that the effect produced by a diode rectifier differs in one important respect from that which would be produced by the ideal rectifier hitherto considered. If the rectifier characteristic were truly a straight line sloping upward from the origin for all positive values, and a straight line of zero slope for all negative values of applied voltage, the charge introduced into the telephone condenser would be directly proportional to the peak value of the applied voltage. Such a rectifier is said to be a linear one. The term linear is often thought to refer to the fact that the characteristic curve is represented by straight lines, but actually it signifies that the output voltage, i.e. e , is directly proportional to the peak signal voltage $\mathcal{V}$ and can be expressed algebraically thus

$$e = K \mathcal{V}$$

where K is a constant of proportion. No such ideal rectifier is known. Referring to the characteristic curve of an actual diode, it is found by trial that for small values of anode voltage, the anode current varies approximately with the square of the anode voltage. This implies that for small applied signal voltages, the diode will give a rectified current which is proportional to the square of the input voltage. Fig. 7 shows the actual characteristic of a diode, and the peak current obtained by applying voltages of various peak values. An alternating E.M.F. of peak value $2\frac{1}{2}$ volts gives a peak current of .5 milliamperes, while 5 volts gives a peak current of 2 milliamperes, that is, doubling the input voltage gives four times the rectified current. Similarly, $7\frac{1}{2}$ volts input gives a current of 4.5 milliamperes peak value. If however, the input voltage is further increased this "square law" is not maintained, for an E.M.F. of 10 volts peak only causes a current of 7.5 milliamperes peak value, whereas if the square law were applicable the current would be 8 milliamperes, while a further increase of input voltage to 12.5 volts gives only 11 milliamperes instead of 12.5 milliamperes which would be obtained if the square law continued to represent this portion of the characteristic. These effects are generally expressed by the statement that the characteristic of a diode rectifier follows the square law ($I_a \propto V_a^2$) over the lower portion, and the linear law ($I_a \propto V_a$) over the upper

CHAPTER X.—PARA. 10

portion. There is no hard and fast line of demarcation, but a few trial values as given above will decide in what region either law may be assumed to hold. This law of variation between input voltage and resulting current has been explained with specific reference to the diode, but it is found that practically all rectifiers in general use possess such characteristics that for input voltages below a certain value the square law may be assumed, while for higher input voltages the linear law is more nearly correct.

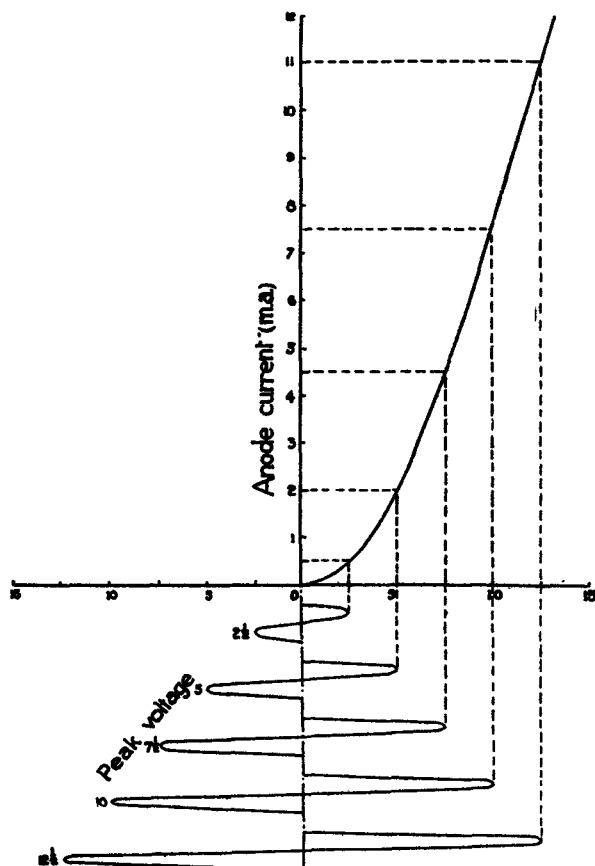


FIG. 7, CHAP. X.—Relation between peak input voltage and peak anode current in actual diode.

10. In the foregoing discussion it was assumed that in the absence of a signal the anode and filament of the diode were at the same potential. This condition is not necessary and in practice it is generally found advantageous to apply a small positive potential to the anode of the valve in order to obtain the greatest rectified current for a given signal voltage. The rectification of a single cycle is then as shown in fig. 8. Under no-signal conditions, the anode voltage is E_a and the anode current I_0 . During the positive half-cycle the anode current gradually increases to I_1 and falls to I_0 , while during the negative half-cycle it falls to I_2 , returning to the value I_0 at the end of the cycle. As the increase of anode current ($I_1 - I_0$) is greater than the decrease ($I_0 - I_2$) the average current during the cycle is not I_0 but I_m , where

$$I_m = \frac{I_1 + I_2}{4} + \frac{I_0}{2}$$

In the diagram, $I_1 = 4.5$ milliamperes, $I_2 = 0.5$ milliamperes and $I_0 = 2$ milliamperes, hence

$$I_m = \frac{4.5 + 0.5}{4} + 2$$

$$= 2.25 \text{ milliamperes.}$$

11. The increase of anode current due to the application of a signal voltage is the rectified current. It is equal to $I_m - I_0$ or to $\frac{I_1 + I_2}{4} - \frac{I_0}{2}$; in the present example this is equal to 0.25 milliamperes. In fig. 8 this result has also been derived graphically. The construction is as follows:—Join I_1, I_2 by a straight line, intersecting the voltage ordinate E_a at I_3 . Bisect the straight line I_0, I_3 . The bisecting line corresponds to the current ordinate I_m , and the rectified current is the difference between I_m and I_0 as previously stated. The point I_0 on the

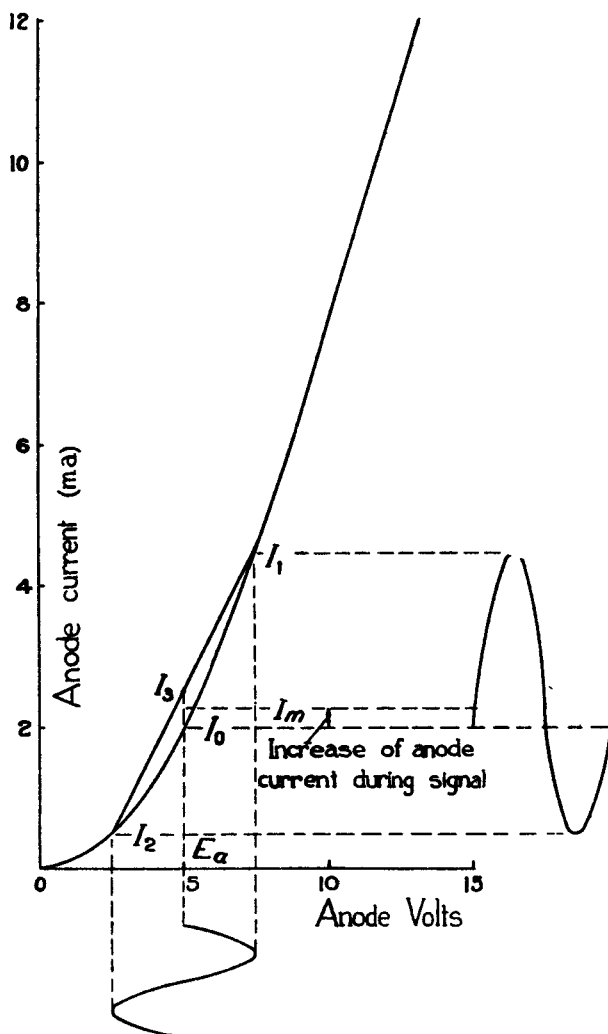


FIG. 8, CHAP. X.—Rectified current, square law rectification.

CHAPTER X.—PARA. 12

characteristic curve, which gives the anode current under no-signal conditions, is called the operating point and it can be proved that to obtain maximum rectified current the operating point should be located where the curvature of the characteristic is greatest. In this connection curvature may be explained as follows. Any characteristic curve may be considered to be built up of a series of arcs of circles, these arcs having different radii and the arcs themselves being very small. Then the arc having the smallest radius is the one having the greatest curvature, and the operating point should be situated at the midpoint of this arc. In many practical cases, the position of this point can be determined approximately by inspection of the characteristic. This rule, namely that the operating point should be located at the point of greatest curvature, is applicable to all rectifiers when the applied E.M.F. is small, i.e. less than about one volt.

C.W. RECEPTION

12. It has already been shown that if the radio-frequency voltage applied to the rectifier is of constant amplitude, as it would be during the reception of continuous wave signals, the output of the rectifier will take the form of a direct current, and will cause no vibration of the telephone diaphragm. A simple expedient which achieves the desired result is to introduce some form of interrupter in series with the telephones, preferably operated by an other than electrical means (fig. 9). The operation of this device is as follows. In the absence of a signal E.M.F. no anode current flows through the telephone windings and no sound is produced, in spite of the fact that the interrupter is in operation. When a signal is received, the voltage applied to the rectifier introduces a charge into the reservoir condenser in the manner previously

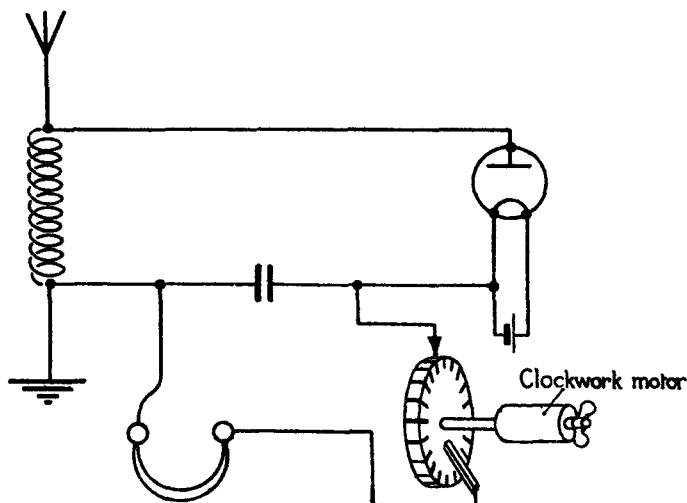


FIG. 9, CHAP. X.—Principle of tikker reception of C.W. signal.

described, and a discharge occurs whenever the telephone circuit is completed by the interrupter. The current through the telephone windings consists of a series of unidirectional pulses and the diaphragm is thrown into vibration at a rate corresponding with the number of interruptions per second; a device of this nature is called a "tikker". This was the first practical method of C.W. reception, and in an emergency might still prove to be of value, but it has been entirely supplanted by a much more effective method which is known as beat reception or heterodyne reception.

Heterodyne reception

13. The importance of the phenomenon known as the "beat effect" has for long been recognised in the study of sound. An example familiar to every airman is found in the sound produced by a twin-engined aeroplane when the engine speeds differ slightly. If one engine is running at 2,000 r.p.m. while the other is running at 2,060 r.p.m. the combined sound of the engines will rise and fall in intensity once per second, and this variation in loudness is termed the beat effect. As the vibrations in this example are very complex, consider two bodies which when set into vibration will set up sinusoidal sound waves in the surrounding air, for example two tuning forks which have been individually adjusted to emit a sound having the pitch of first lower C, corresponding to a frequency of 128 cycles per second. If both are sounded together, the aural impression gained by an observer is that of a single clear note, but if a small piece of wax is attached to each prong of one fork, so that its mass is increased, this fork will now have a natural frequency slightly lower than the other, e.g. 126 cycles per second. On sounding both

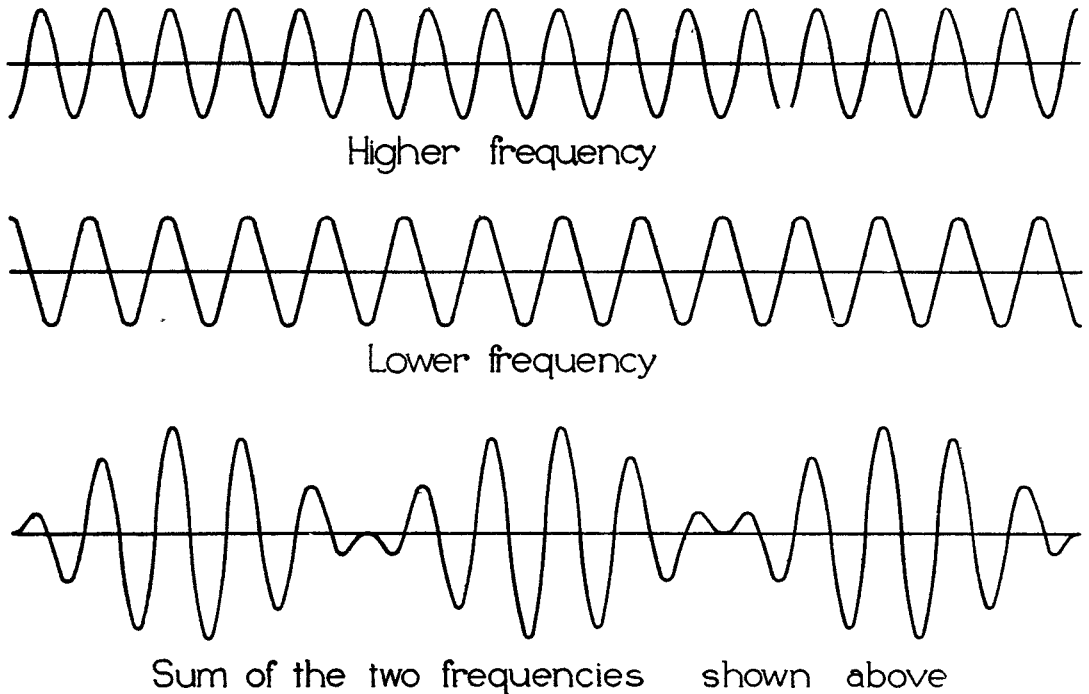


FIG. 10, CHAP. X.—Beat effect between waves of different frequencies and equal amplitude.

simultaneously, it will be observed that the emitted sound waxes and wanes in intensity twice per second; the explanation is as follows. The intensity of the sound at any instant is the sum of the intensities due to the two separate sources. The sound waves are both sinusoidal in form but of different frequency, and this instantaneous sum can be found by plotting the two curves showing the instantaneous intensity due to each source over a period of one or more seconds, subsequently finding the sum of the two curves at a large number of points on the "time" axis, and plotting this sum to form a new curve. This process has been performed in fig. 10 in which are shown the two component sine curves, the latter being of equal amplitude in this particular example. The curve obtained by adding the instantaneous values of each curve is also shown, this curve being called the resultant of the other two. The individual half-cycles of this resultant are approximately sinusoidal, but the amplitude of each successive half-cycle varies in value, growing from zero to a maximum, falling again to zero, again rising to a maximum

CHAPTER X.—PARAS. 14-15

and so on, so long as the two sources continue to produce their characteristic sound. The maximum amplitude of the resultant is twice the amplitude of either component curve and the number of beats per second is equal to the difference between the two component frequencies.

14. The laborious process of graphical addition can be replaced by mathematical addition of the two equations representing the component sine waves. Let us assume that these equations are

$$v_1 = \mathcal{V} \sin \omega_1 t, \quad \omega_1 = 2\pi f_1$$

$$v_2 = \mathcal{V} \sin \omega_2 t, \quad \omega_2 = 2\pi f_2$$

$(f_1 - f_2)$ being small compared with f_1 and f_2 . It will be convenient to denote the mean frequency $\frac{f_1 + f_2}{2}$ by f , and $2\pi f$ by ω ; then $\omega = \frac{\omega_1 + \omega_2}{2}$. The equation representing the resultant wave will then be

$$v = v_1 + v_2 \text{ or}$$

$$v = \mathcal{V} (\sin \omega_1 t + \sin \omega_2 t)$$

The sum of the sines of two angles, say P and Q , is given by the equation

$$\sin P + \sin Q = 2 \sin \frac{P+Q}{2} \cos \frac{P-Q}{2}$$

Hence

$$\begin{aligned} v &= 2\mathcal{V} \sin \frac{\omega_1 + \omega_2}{2} t \cos \frac{\omega_1 - \omega_2}{2} t \\ &= 2\mathcal{V} \cos \frac{\omega_1 - \omega_2}{2} t \left\{ \sin \omega t. \right. \end{aligned}$$

The expression has been bracketed in such a way as to emphasise that the resultant is a sinusoidal quantity of frequency $\frac{f_1 + f_2}{2} = f$, the maximum amplitude of which is $2\mathcal{V}$. As shown in fig. 10, the amplitude varies between zero and $2\mathcal{V}$, a complete cycle of this variation occurring $f_1 - f_2$ times per second. The number $f_1 - f_2$ is often termed the "beat frequency," but it is desirable to avoid the use of the term "frequency" when referring to the number of beats per second, reserving its employment to express the number of cycles per second of an alternating quantity.

15. When the amplitudes of the two component vibrations are not equal, the resultant is of somewhat more complicated form, in that the periodic time of the individual waves composing a complete beat varies slightly during each beat, while the amplitude of the vibration varies between two values which are the sum and difference of the amplitudes of the constituent quantities, e.g. if the latter are represented by A and B , the amplitude of the resultant varies between $A + B$ and $A - B$. The resultant beats caused by the super-position of two vibrations of different amplitude is shown in fig. 11. Prior to the rectification of a C.W. signal it is necessary to cause an audio-frequency variation in amplitude of the voltage which is to be rectified, in order to produce an audio-frequency change in the telephone current, and the phenomenon just described is easily adapted to this end. The method is generally referred to as "heterodyne reception," the term heterodyne having been coined to denote the "mixing" of power from different sources. There are two methods of applying the heterodyne principle, which are termed (i) "separate heterodyne" and (ii) "autodyne" or "self-heterodyne." The latter cannot be used with a diode detector, and will be explained later.

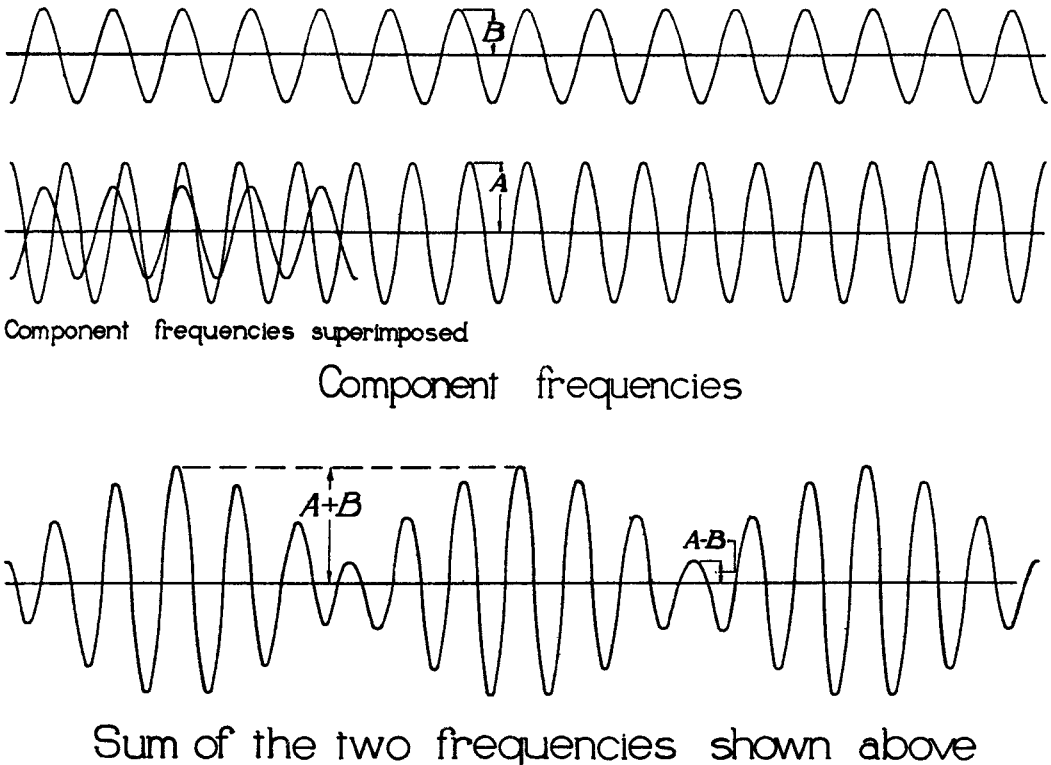


FIG. 11, CHAP. X.—Beat effect between waves of different frequencies and unequal amplitude.

The separate heterodyne

16. In this system we may consider the receiver proper to consist of the tuned aerial circuit, the diode and its ancillary battery, and the telephone receiver. We have seen that this apparatus alone will not cause the telephone diaphragm to vibrate at an audible frequency on the arrival of a C.W. signal. In separate heterodyne reception, a small oscillator of extremely low power is located near to the receiving aerial circuit. This oscillator generates a radio-frequency current in an oscillatory circuit which is tuned to within about 2,000 cycles per second of the signal to be received, and is coupled to the receiver aerial system. A radio-frequency E.M.F. of the local oscillator frequency is consequently set up in the aerial circuit, but this E.M.F. alone will produce no variation of telephone current, for reasons already explained. On the receipt of a C.W. signal, however, an E.M.F. of the signal frequency is also induced in the aerial circuit, and the oscillatory current in the latter is the sum of two currents of different frequencies, i.e. those of the C.W. signal and local oscillator respectively. Beats of the nature shown in fig. 10 or 11 are therefore produced, and the voltage induced in the aerial inductance has a similar waveform. This voltage is applied to the anode-filament path of the diode and its rectification will result in a pulsating current in the anode circuit and an audio-frequency variation of telephone current, hence the action upon the diaphragm is as shewn in fig. 12. The simplest form of receiver using the separate heterodyne system is given in fig. 13; the receiver proper is similar to that given in fig. 4, and the circuit of the local oscillator identical with that of the simple oscillator described in Chapter IX. The aerial circuit is tuned to the frequency of the distant transmitter by means of the inductance L and condenser C , and the local oscillator to some frequency differing from that to be received by from 200 to 2,000 cycles per second. The local oscillator is said to be detuned from resonance by this amount, and the degree of detuning is

CHAPTER X.—PARA. 16

often of great assistance in achieving freedom from interference due to the operation of other transmitters. To understand this, imagine the transmitter from which reception is desired to emit a continuous "dash," and the receiving aerial circuit to be correctly tuned for its reception. If the local oscillator is adjusted to generate a frequency 20,000 cycles above the desired frequency, the interaction of the two frequencies will give rise to 20,000 beats per second; although the detector valve will operate in such a manner that the current flowing in the telephones undergoes a practically sinusoidal variation in amplitude at the same rate, no appreciable sound is perceived owing to the small response of the telephone diaphragm and the insensitivity of the ear to such a high-pitched note. As the frequency of the local oscillation is brought near that of the distant transmitter the number of beats per second is reduced, for example, when the frequency difference is about 5,000, a very high note of feeble intensity will be perceived. This note is too highly pitched and of too small amplitude to be suitable for morse code operation, and is generally said

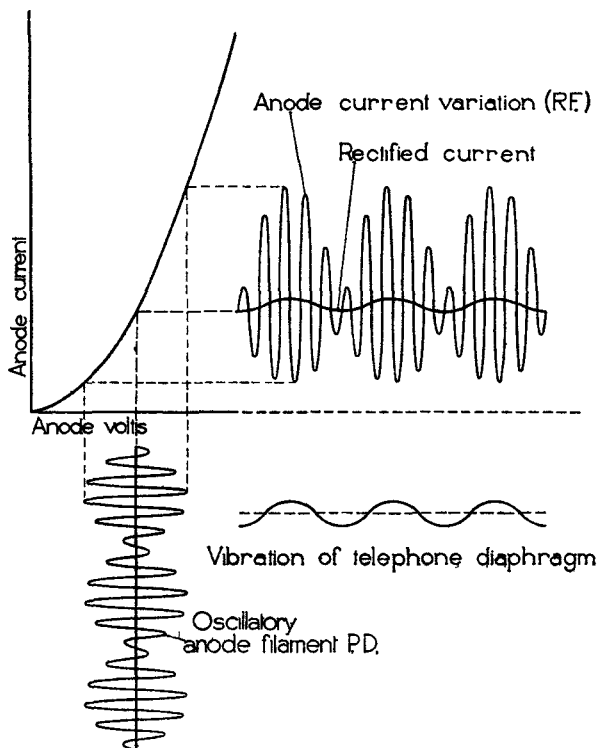


FIG. 12, CHAP. X.—Rectification of heterodyne beat voltage by diode.

to have "no body." On bringing the local oscillator more nearly into resonance with the signal frequency, e.g. differing only by 2,000 cycles per second, the note becomes lower pitched and possesses more "body" because the telephone diaphragm vibrates with greater amplitude and the ear is more sensitive to sound of this frequency. If the local oscillator is left at this point and actual reception of morse code performed for a few minutes, this note will be found tiring to the ear. A further approach to resonance between local oscillator and distant transmitter will give a lower-pitched note with (apparently) still more sound energy; most operators usually prefer a note which is in the neighbourhood of 800 to 1,000 cycles per second, although of course the setting is done entirely without reference to any scale either of frequency of local oscillation or pitch of note. The operator simply varies the adjustment of the local oscillator (usually the capacitance of the local oscillatory circuit) until the note suits him.

Dead space

17. If the frequency of the local oscillator is brought within about 200 cycles per second of the transmitter frequency, the beat effect generally ceases. This is due to two causes, first, the insensitivity of the telephone receiver at low audio-frequencies, and second, a phenomenon

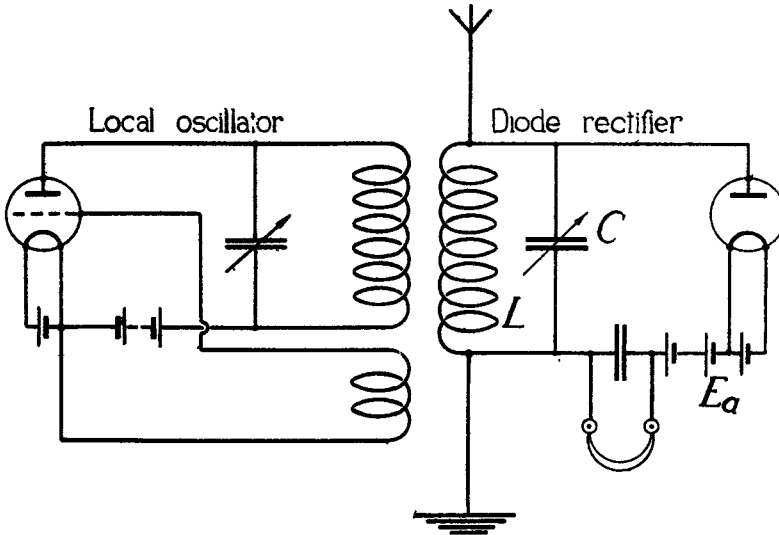


FIG. 13, CHAP. X.—Separate heterodyne receiver using diode detection.

called automatic synchronization. When the tuned circuit of the local oscillator becomes nearly resonant with the distant transmitter, it is forced to oscillate at the frequency of the transmitter instead of at its own natural frequency, and under this condition no heterodyne beats are produced. Under practical conditions it will be found that this usually occurs over a

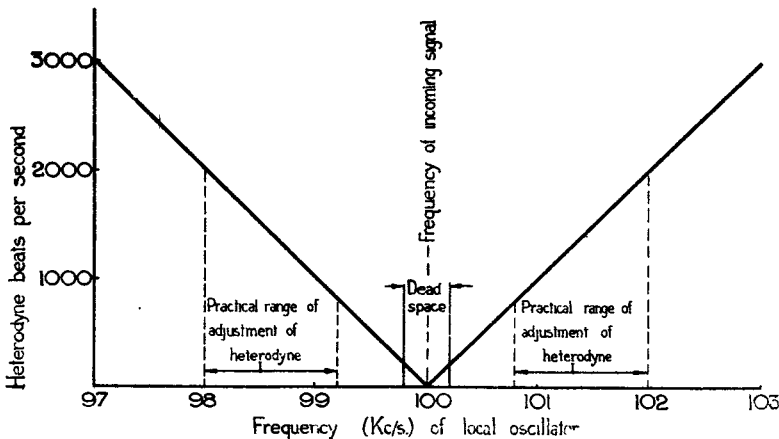


FIG. 14, CHAP. X.—Variation of beat note with frequency of local oscillation.

frequency band of about 200 cycles per second above and below the frequency of the transmitter, and this zone is known as the "dead space" of the heterodyne. If the adjustment of the heterodyne is continued through the dead space, the beat effect is again perceived, first giving a very low pitched note having a frequency in the region of 200 cycles per second, and increasing

CHAPTER X.—PARAS. 18-19

to a pitch above the limit of audibility as the frequency difference is increased. The change of pitch of the heterodyne beat-note is illustrated in fig. 14 for a particular instance, in which the incoming signal is assumed to have a frequency of 10^6 cycles per second, and the frequency of the local oscillator is continuously varied between 97,000 and 103,000 cycles per second. The straight sloping lines represent the number of beats per second which are formed by the interaction of the two frequencies, while the dotted lines show the practical limits between which the frequency of the local oscillator can be varied, observing the limitations caused by the lack of sensitivity of the ear and telephone diaphragm at frequencies much over 2,000 cycles per second.

Advantages of heterodyne reception

18. The fact that the pitch of the telephone signals is entirely under the control of the operator is one of the advantages of C.W. over I.C.W. operation, the note in the latter case being fixed by apparatus at the transmitter. In addition to allowing (under normal conditions) the use of a pitch which is agreeable to the ear of each particular operator, thus enabling him to handle traffic with less fatigue, it gives a certain degree of what may be termed "note selectivity." Thus the operator receiving a signal on a frequency of 10^6 cycles per second may experience interference on a frequency adjacent to this, say 1,001,000 cycles per second. If the operator sets his local oscillator to 998,500 cycles per second, the desired signal will, after rectification, give a note of frequency 1,000,000-998,500 or 1,500 cycles per second, a rather high note not very difficult to read, in fact one which may be preferred by some operators. The interfering signal will, after rectification, give a note of 1,001,000 - 998,500 = 2,500 cycles per second, which is of higher pitch and will also be of reduced strength owing to the comparatively inferior response of the telephone at this frequency. Alternatively, the operator may adjust his local oscillator to within 200 cycles per second of the interfering signal, which then falls in the zone referred to above as the dead space; no beat note is then caused by the interfering signal, while the desired signal gives a beat note of between about 800 and 1,200 cycles per second according to the exact point tuned to in the dead space. For this expedient to be entirely successful, it is necessary for the transmitter frequencies to be controlled in such a manner that a high degree of constancy is maintained. A further advantage of heterodyne reception is that if the rectifier obeys the square law ($I_a \propto V_a^2$) as is always the case for small input voltage, the rectified current is not proportional to the signal voltage but to the strength of the local oscillation, hence a C.W. signal received in this way will give a much louder response than an I.C.W. or tonic train signal. This can be shown as follows.

19. Let the input voltage of the signal be $v_1 = \mathcal{V}_1 \sin \omega_1 t$ and the local oscillator voltage be $v_2 = \mathcal{V}_2 \sin \omega_2 t$. The total input voltage will then be $v = v_1 + v_2$ or

$$v = \mathcal{V}_1 \sin \omega_1 t + \mathcal{V}_2 \sin \omega_2 t.$$

The characteristic of the rectifier is given by the equation

$$i_a = K v_a^2$$

Therefore

$$\begin{aligned} i_a &= K \{ \mathcal{V}_1^2 \sin^2 \omega_1 t + \mathcal{V}_2^2 \sin^2 \omega_2 t \} \\ &= K \{ \mathcal{V}_1^2 \sin^2 \omega_1 t + 2 \mathcal{V}_1 \mathcal{V}_2 \sin \omega_1 t \sin \omega_2 t + \mathcal{V}_2^2 \sin^2 \omega_2 t \} \end{aligned}$$

The expression in braces may be divided into two groups of terms:—

$$\begin{aligned} \text{(i)} \quad & \mathcal{V}_1^2 \sin^2 \omega_1 t + \mathcal{V}_2^2 \sin^2 \omega_2 t \\ &= \frac{\mathcal{V}_1^2}{2} + \frac{\mathcal{V}_1^2}{2} - \frac{\mathcal{V}_1^2}{2} \cos 2 \omega_1 t - \frac{\mathcal{V}_2^2}{2} \cos 2 \omega_2 t. \end{aligned}$$

These terms cause a mean rise of anode current, and also radio-frequency variations of twice the frequency of the incoming and local oscillation respectively; these do not affect the telephone windings but are by-passed by the reservoir condenser.

$$\begin{aligned} \text{(ii)} \quad & 2 \mathcal{V}_1 \mathcal{V}_2 \sin \omega_1 t \sin \omega_2 t \\ &= \mathcal{V}_1 \mathcal{V}_2 \{ \cos (\omega_1 - \omega_2) t - \cos (\omega_1 + \omega_2) t \} \end{aligned}$$

The second term within the braces is also a radio frequency having no effect on the telephone windings, but the first corresponds to the number of heterodyne beats per second, and will result in an audio-frequency component of anode current, its amplitude being $K\mathcal{V}_1\mathcal{V}_2$. The strength of the signal is therefore not proportional to $\mathcal{V}_1$ but to the product $\mathcal{V}_1\mathcal{V}_2$. The amplitude of the signal voltage, i.e. $\mathcal{V}_1$, is not under control, but that of the local oscillator, i.e. $\mathcal{V}_2$, is variable within wide limits by adjustment of the coupling between local oscillator and receiver proper. The signal strength due to the voltage $\mathcal{V}_1$ may therefore be increased by an increase in the voltage $\mathcal{V}_2$. The notion somewhat prevalent among wireless operators that optimum conditions are obtained when the signal and local oscillator voltages are equal, is thus shown to be fallacious; the amplitude of the local oscillation may profitably be increased provided that the characteristic follows the law $i_a = Kv_a^2$ over the operating range, but a further increase in the amplitude of local oscillation will not result in a louder response from the telephones. In certain circumstances such an increase in the amplitude of the local oscillation will cause the signal to become weaker.

THE TRIODE RECTIFIER

20. Although the principles of rectification exemplified by the diode are of general application, this type of valve is not in common use as a rectifier in receivers designed to deal with very small input voltages. It has however, a considerable field of employment in broadcast reception, in which the signal voltage is considerably amplified before rectification. (See Chapters XI and XII.) Where it is desired to obtain the strongest signal in the telephone receivers for a given input voltage, the triode valve is almost universally employed. It has previously been stated that this type of valve functions primarily as a voltage amplifier, and when operated under conditions in which rectification is also obtained, a more powerful response can be obtained than by the process of rectification alone. As before, it is desirable to consider the reception of I.C.W. or tonic train signals in the first place, in order to avoid the complications caused by the necessity for beat reception of C.W. The triode is usually employed as a rectifier either by making use of (i) the curvature of the grid current/grid volts curve, giving rise to what is termed cumulative grid rectification, or (ii) the curvature of the anode current/grid volts curve,

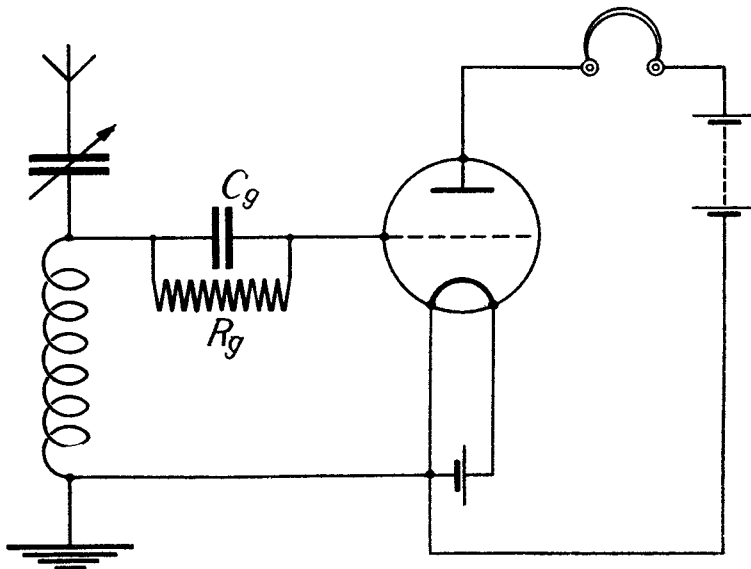


FIG. 15, CHAP. X.—Cumulative grid rectification.

CHAPTER X.—PARAS. 21-22

a method generally known as anode bend rectification. Each method has several alternative methods of circuit connection and operating conditions, but these do not affect the general principles.

Cumulative grid rectification

21. The simplest form of circuit for use with this form of rectification is given in fig. 15. The aerial circuit is tuned to the frequency of the desired signal as usual, and the voltage developed in the inductance is applied to the grid and filament of the triode. In series between the inductance and the grid of the valve is interposed a condenser C_g , called the grid condenser, the reactance of which is negligible at the signal frequency. In parallel with this is a resistance R_g , called the grid leak, which may be one or more megohms; the factors which determine the values of the capacitance C_g and resistance R_g will receive consideration later. The anode circuit of

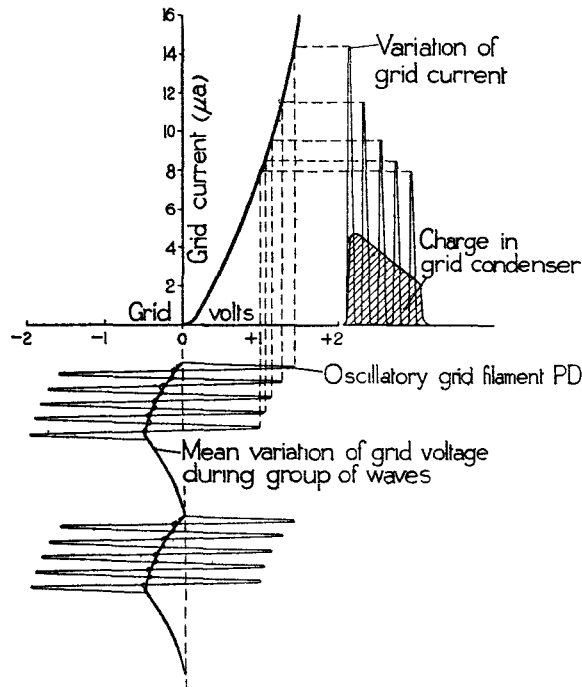


FIG. 16, CHAP. X.—Action in grid circuit of cumulative grid rectifier.

the valve consists of the anode-filament path of the latter, a pair of telephones and the anode H.T. battery; in discussing the initial stages of the action we may ignore the anode circuit altogether and confine our attention to the grid circuit. The presence of a signal E.M.F. will set up in the aerial circuit an oscillatory current of the same frequency, and consequently a counter-E.M.F. is developed in the inductance L . This E.M.F. is applied to the circuit consisting of the grid condenser (with its leak resistance) in series with the grid-filament path of the valve, and as the impedance of the latter path is very large compared with that of the parallel combination of grid leak and condenser, practically the whole of the oscillatory voltage is applied between grid and filament of the valve.

22. In order to predict the current which will flow under the influence of this voltage, the characteristic grid current/grid volts curve of the particular valve must be consulted. Fig. 16

shows the $I_g - V_g$ curve of a suitable receiving valve, and it will be observed that when the grid and filament are at the same potential, no electron current flows from filament to grid. Below the $I_g - V_g$ curve are shown the grid-filament voltage variations corresponding to two groups of waves which are assumed to be part of, say, a morse "dot," as in the explanation of the diode rectifier. When the grid is positive with respect to the filament a grid current will be established, but the grid-filament path is perfectly non-conductive during those half-cycles of grid voltage in which the grid is negative to the filament. On the first positive half-cycle therefore, electrons will be attracted by the grid, and will pass into the right hand side of the grid condenser charging it negatively, and causing a displacement current in the dielectric with the consequent repulsion of electrons from the left hand plate, which thereupon becomes positively charged. During the succeeding half-cycle, in which the grid becomes negative to the filament, no electrons are attracted to the grid, but the grid condenser becomes partly discharged owing to the presence of the grid leak resistance R_g . At the end of the first complete cycle the grid potential is

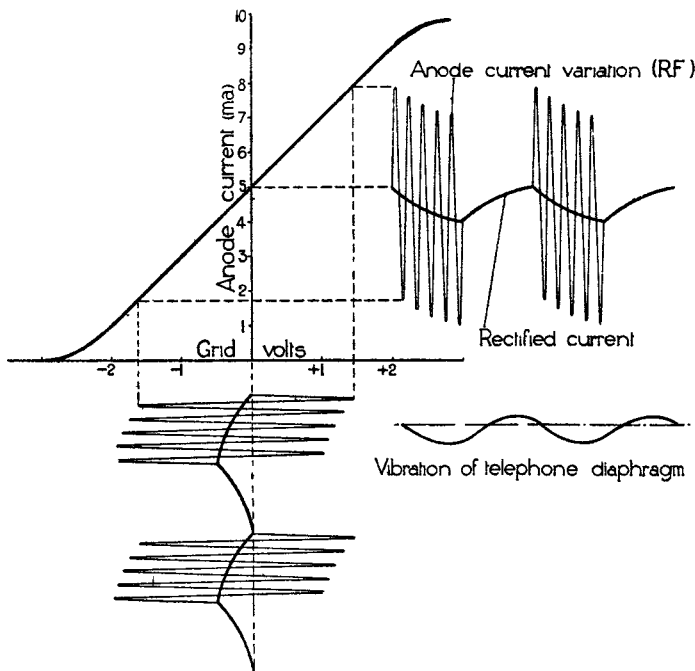


FIG. 17, CHAP. X.—Action in anode circuit of cumulative grid rectifier.

slightly negative compared with its potential at the beginning of the cycle. During the next positive half-cycle, the grid condenser will again receive a negative charge on its right hand plate, and will lose a portion of this charge during the negative half-cycle owing to the conductive property of the resistance R_g ; the net effect is that at the end of the second cycle, the grid has acquired a still further negative potential with respect to the filament, and so on until the end of the group of waves. When no further oscillatory voltage is applied between grid and filament the grid condenser continues to maintain a current through the grid leak until its charge has been completely dissipated, and during this time the grid gradually returns to its original potential, namely, that of the filament. This sequence of operations is shown graphically in fig. 16. The group of waves has thus caused the introduction of a charge into the grid condenser C_g in exactly the same manner as into the reservoir condenser in diode rectification, the resistance R_g acting as the load resistance. If the latter were absent, the grid condenser would retain its

CHAPTER X.—PARAS. 23-24

charge at the end of the group of waves, and the succeeding group would cause the grid to become still further negative and so on. In consequence, after the first few groups of waves, the grid would become so negatively charged that the grid current would be reduced to zero and further rectification would become impossible.

23. (i) The effect upon the anode current of this gradual accumulation of charge in the grid-condenser may now be considered, with the aid of the $I_a - V_g$ curve, fig. 17. Before the oscillatory voltage is applied, the anode current has a steady value of 5 milliamperes; but during the group of waves the grid voltage becomes more and more negative and the anode current gradually falls to 4.2 milliamperes, an anode current change of .8 milliamperes. At the end of the group the anode current resumes its original value of 5 milliamperes as the charge in the grid condenser leaks away and the grid voltage returns to normal. This variation of anode current will cause a single variation of flux acting upon the diaphragm and a corresponding inflection of the latter, or more crudely, a single "click." Since however the groups of waves must be assumed to occur about 1,000 times per second, the diaphragm is set into vibration at this rate and gives rise to the emission of a musical note by the telephone receivers. In addition to this audio-frequency variation of anode current, the latter will also vary at the frequency of the applied signal voltage. This variation does not enter into the process of rectification, but nevertheless may be made to perform a useful function, as explained later.

(ii) As in the case of the diode detector previously described, most efficient rectification of weak signal voltages is obtained when the operating point is located where the curvature of the rectifying characteristic is greatest. With most triodes in use at the present time, this necessitates the application of a small positive potential (generally about .2 to .3 volt) to the grid of the valve. This potential can be obtained by connecting the grid leak to the positive terminal of the filament battery instead of to the negative as shown in fig. 18a. In the following explanation the valve is assumed to require a filament battery of 2 volts, such valves being in general service use. It must not be supposed that this method of connecting the grid leak resistance will cause the grid potential to be 2 volts positive to the negative end of the filament, for the fact that the grid is at positive potential implies that grid current will be established, and the grid-filament conductance of the valve will be of some finite value. Let us suppose that this is $\frac{2}{200,000}$ siemens (mho), i.e. that the effective resistance r_g (for direct current) of the grid-filament path inside the valve is 200,000 ohms, while the grid leak resistance is 2 megohms. In addition to supplying the filament heating current, the filament battery then causes a current to flow in the circuit comprised by the grid-filament path and the grid leak. In the present circumstances this

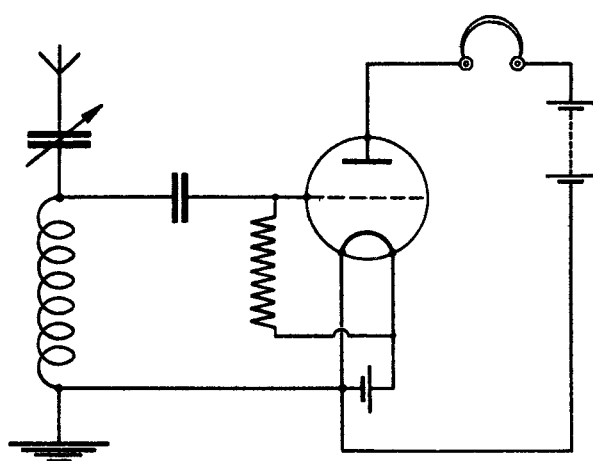
current will be $\frac{2 \text{ volts}}{(2,000,000 + 200,000) \text{ ohms}}$ or .91 microamperes. As a result of this current a potential difference will be established between the grid and filament, and the grid will be positive with respect to the negative end of the filament by an amount equal to $r_g I_g = 200,000 \times \frac{2}{2.2 \times 10^6} = \frac{.4}{2.2}$ or .18 volts.

24. The reader may reasonably object that this value is obtained by an unwarranted assumption as to the value of r_g , for in the $I_g - V_g$ characteristic of fig. 16 measurement of the slope of the curve shews that r_g varies from infinity to about 2×10^5 ohms as the grid voltage varies from zero to $+.18$ volt. For any value of grid leak resistance R_g , and any valve for which the $I_g - V_g$ characteristic is given, however, the operating point can be located by a graphical method from the above principles, for the resistance R_g can be treated as a "load line" (see Chapter VIII) and drawn upon the characteristic. To explain the process, consider the $I_g - V_g$ curve of fig. 18b, and let $R_g = 2 \times 10^6$ ohms; this resistance may be represented by a line having a slope of .5 microampere per volt. With the connections made as in fig. 18a, with no current flowing in the grid leak, there will be no IR drop in the latter and the grid will be 2 volts positive to the filament. Hence one end of the load line will rest at the point $V_g = +2$, $I_g = 0$. If 1 microampere were to flow in the grid leak, the IR drop in the latter would be 2 volts, and the grid potential would be $2 - 2$ or zero volts, hence another point on the load line is that corresponding to $V_g = 0$, $I_g = 1$ microampere, and the line may be drawn with a

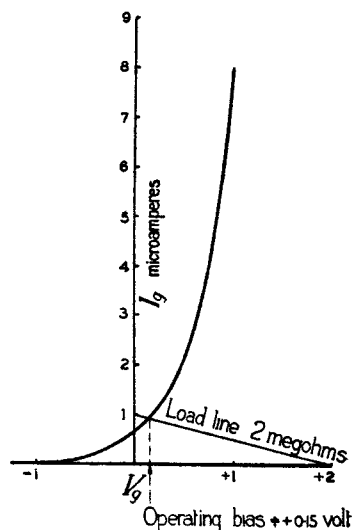
straightedge to connect the two points so found. This line intersects the $I_g - V_g$ curve at a point at which the grid current through the valve is equal to the current through the leak, and the point of intersection gives (i) the mean value of grid current when no signal E.M.F. is applied and (ii) the mean potential of the grid under these conditions.

Damping due to grid current

25. Having stated that the application of a small positive potential results in greater rectified current for a given grid-filament voltage, it may appear paradoxical to state that the effect of such positive grid bias may be to reduce the strength of signals. That this is so will be appreciated when it is remembered that the E.M.F. applied between grid and filament depends partly upon the opposition offered by the path. In the absence of grid current, the path may be considered to consist only of a very small capacitance having no losses and consequently causing



(a)



(b)

FIG. 18, CHAP. X.—Alternative connection of grid leak resistance.

no waste of energy. When electrons flow from filament to grid, passing round the external circuit back to the filament, the electrons inside the valve can be urged into oscillatory motion by an applied alternating E.M.F. just as if they were enclosed in a conductor. Hence the impedance of the grid-filament path, instead of being purely due to the small inter-electrode capacitance, consists of this capacitance in parallel with a resistance r'_g —the A.C. resistance of the grid-filament path. The shunting effect of the resistance r'_g reduces the circuit magnification and therefore, for a given induced E.M.F. in the aerial, the voltage actually applied to the grid-filament path is reduced. This effect will receive further consideration in Chapter XI.

Factors influencing the values of grid condenser and leak

26. The values of capacitance and resistance chosen for these components is of great importance. Attention must be given to four considerations. First the time constant of the circuit (consisting of grid condenser and the parallel combination of grid leak resistance R_g and the internal grid-filament path r_g) should be of such a value that the charge in the condenser has time to leak away in the interval between consecutive groups of waves. Second, the capacitance of the grid condenser must be small in order that a small accumulation of electrons

CHAPTER X.—PARA. 27

may charge it to as high a potential as possible. Third, the reactance of the grid condenser at the signal frequency should be small compared with the impedance of the grid-filament path of the valve, in order that the voltage drop between the condenser terminals shall not be an appreciable fraction of the signal voltage. This consideration points to a large value of C_g and is directly opposed to the second requirement. Fourth, the resistance of the grid leak should be as large as possible in order that the accumulation of electrons on the grid condenser may build up a considerable potential between the terminals of the latter; if too large a value is adopted, however, the first requirement will not be fulfilled. In addition to these four requirements for maximum signal strength, the question of selectivity is also involved. The large number of variable quantities renders it advisable to settle the capacitance of the grid condenser and resistance of the grid leak by direct experiment, bearing in mind the service for which any particular receiver is designed; in most cases a capacitance of from $\cdot 0001$ to $\cdot 0003 \mu F$ and a resistance of from $\cdot 5$ to 2 megohms are found suitable.

Damping due to anode circuit load

27. The effect of grid current in decreasing the magnification of the input circuit of the triode has already been pointed out. It is perhaps not quite so obvious that, even if no grid current flows, the valve and its associated anode circuit will impose a load upon the effective voltage of the tuned circuit. This phenomenon is generally called the Miller effect after the physicist who

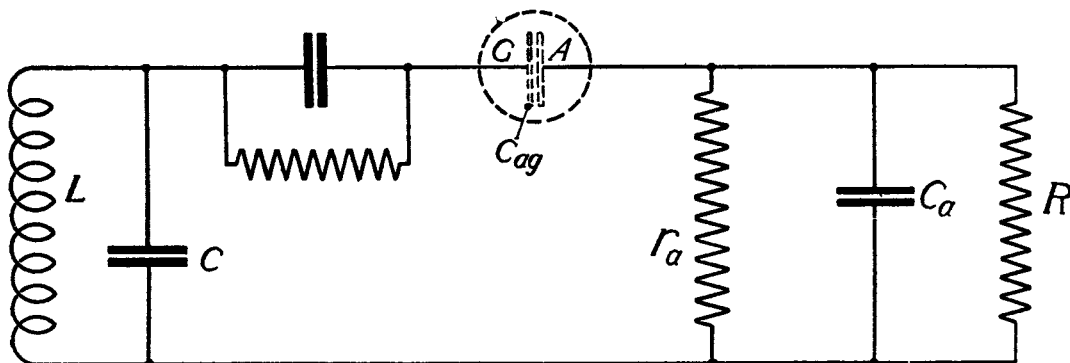


FIG. 19, CHAP. X.—Damping of input circuit owing to anode load and grid-anode capacitance.

first analysed its cause. The Miller effect is due to the electrostatic coupling which exists between the anode and grid circuit. This coupling may be, and in practice generally is, only that caused by the capacitance between the grid and anode of the valve, including of course the leading-in wires at the "pinch", which is of the order of $10 \mu\mu F$ in ordinary triodes. The effect is illustrated in fig. 19, in which the anode circuit is assumed to consist of a resistance R and capacitance C_a in parallel, and is effectively coupled to the tuned circuit LC by the grid-anode capacitance C_{ag} . It has already been shown that when circuits are coupled together by a common reactance, an effective resistance is transferred from the secondary to the primary, and this is so in the circuit shown, but the effect is more complex than is apparent from the diagram owing to the amplifying properties of the valve. Further consideration of the Miller effect will therefore be found in Chapter XI. For the present it is sufficient to realise that in extreme cases, the total damping imposed upon the tuned circuit by the valve, including both the effects of grid current and the anode circuit load, may be equivalent to the insertion into the tuned circuit of a series resistance of the order of several hundred ohms. The damping effect of the anode load can be reduced by the employment of a telephone condenser of large capacitance, but this solution leads to a reduction of telephone current and consequently to weaker signals. A more practical expedient will now be discussed.

Reduction of damping by regenerative amplification

28. When no signal voltage is applied between grid and filament of the rectifying valve, the anode current is constant, but during the reception of an I.C.W. or T.T. wave group, it varies in a complex manner. In the simplest instance this complex variation may be considered to consist of the sum of two variations superimposed upon the steady or no-signal anode current, viz. a variation which takes place at a rate corresponding to the number of groups radiated per second, and a variation at the frequency of the applied E.M.F., i.e. the true signal frequency. It has already been stated that the latter component plays no part in the mechanism of rectification proper, and it is now proposed to explain how this component can be beneficially employed. When the grid and anode circuits of a triode are inter-coupled, any variation of grid voltage, by causing a change of anode current, will cause a corresponding E.M.F. to be induced in the grid circuit. Let us therefore suppose that the receiving circuit has been arranged as shown in fig. 20 which is similar to fig. 18a except that an inductance L_a has been inserted in series between the anode of the valve and the telephones. This coil is inductively coupled to the inductance L in the aerial circuit, as indicated by the arrow linking the two coils. Now suppose that the actual

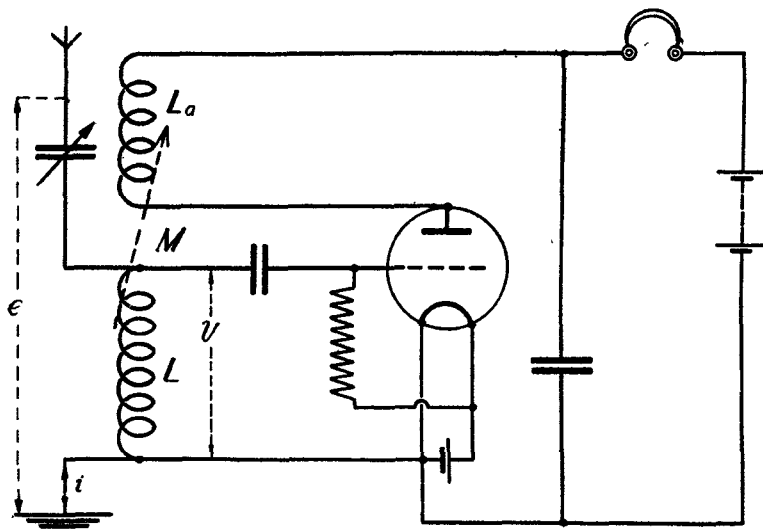


FIG. 20, CHAP. X.—Cumulative grid rectifier with reaction.

E.M.F. induced in the aerial (indicated by ϵ in the diagram) causes grid-filament voltage variations as shown in fig. 17; the resulting anode current variations will then also be the same as in this diagram, and will flow through the coil L_a . We are now only concerned with the radio-frequency component shown by the thinner line. Owing to the coupling between the two coils, the radio-frequency variation of anode current will set up a corresponding flux in and around the coil L_a which links with the coil L , and an oscillatory E.M.F. of the same frequency as the original signal, will be induced in the aerial circuit. The oscillatory grid-filament P.D., v will now be the vector sum of the original P.D. and that due to the coupling between anode and grid; if they are very nearly in phase with each other, v will be equal to their sum and the rectified current will be much greater than in the absence of the coupling arrangement. On the other hand if the two components of the grid-filament P.D. are nearly 180° out of phase, v will be approximately equal to their difference and the rectified current will be smaller than if the signal E.M.F. alone were operative. The coil L_a is referred to as the reaction coil, and the process by which an increase of grid-filament P.D. is achieved as regenerative amplification. The effect of regenerative amplification upon an incoming signal voltage is shown in fig. 21. The upper

CHAPTER X.—PARA. 29

curve shows the grid-filament P.D. due to a tonic train signal without regenerative amplification and the lower curve the grid-filament P.D. with regeneration. It should be noted that in addition to an increase in the amplitude of the successive waves, the oscillation takes longer to reach its maximum amplitude, and dies away more slowly, so that the duration of the wave train is increased. This signifies that the effect of reaction is to reduce the damping (or increase the magnification), of the aerial circuit and therefore, as shown in Chapter V, to increase the selectivity. Instead of stating that the total resistance is decreased, it is convenient to consider that the effect of reaction is to introduce a "negative resistance" into the oscillatory circuit.

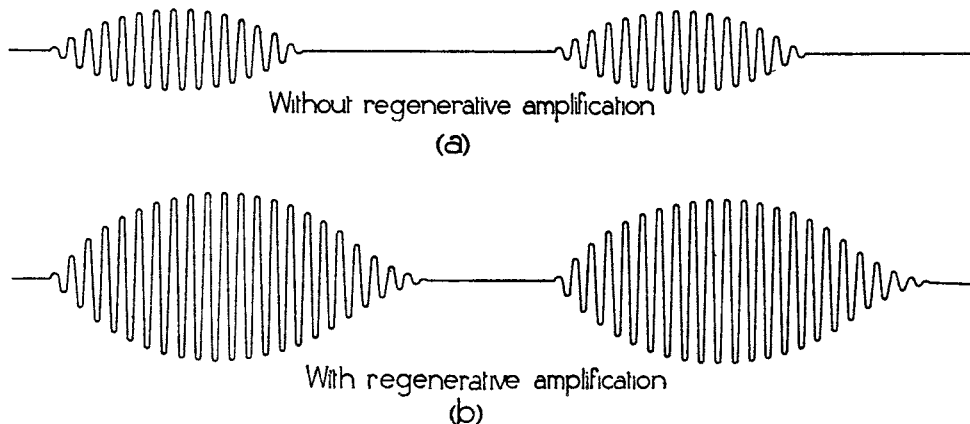


FIG. 21, CHAP. X.—Effect of reaction on grid voltage due to tonic train signal.

Approximate mathematical treatment

29. The effect just explained can be traced out step by step with the aid of fig. 20. The effective voltage set up in the aerial by the distant transmitter is e volts, and we may assume that in conjunction with the E.M.F. due to reaction, this gives rise to a current of i amperes. The counter-E.M.F. set up in the inductance L will be $v = \omega Li$ volts, and this is the signal voltage to be rectified; ignoring the latter process which does not affect the result in any way, this voltage will cause a radio-frequency component of anode current, equal to $\frac{\mu v}{Z_a} = \frac{\mu \omega Li}{Z_a}$ volts, where Z_a is the impedance of the whole anode circuit. The latter consists of the anode A.C. slope resistance r_a , the reactance of the coil L_a and the impedance of the telephones and telephone condenser in parallel. The capacitance of this condenser is purposely adjusted to a value which offers no appreciable reactance at radio frequencies, while the reactance of L_a will be negligible compared to r_a , hence the radio-frequency component of the anode current is approximately equal to $\frac{\mu \omega Li}{r_a}$ amperes. In addition to this, however, by ordinary transformer action an induced E.M.F. is set up in the reaction coil by the aerial current, its value being ωMi volts, consequently the total radio-frequency E.M.F. in the anode circuit is $(\omega M + \mu \omega L) i$ volts and the current caused by the latter is $i_a = \frac{(\omega M + \mu \omega L) i}{r_a}$ amperes. This current reacts into the aerial circuit, setting up an E.M.F. $-\omega Mi_a$ volts, and therefore the total voltage acting in the aerial circuit is not e alone, but $e - \omega Mi_a$ or $e - \frac{\omega M (\omega M + \mu \omega L) i}{r_a}$. Hence if Z is

the impedance of the aerial circuit the aerial current is given by the equation

$$i = \frac{e - \frac{\omega M i (\omega M + \omega L \mu)}{r_a}}{Z}$$

$$\frac{e}{Z} = \left\{ 1 + \frac{\omega M (\omega M + \omega L \mu)}{r_a Z} \right\} i$$

$$\text{and } i = \frac{e}{Z + \frac{\omega M (\omega M + \omega L \mu)}{r_a}}$$

30. The effective impedance added to the aerial by the circuits connected thereto is therefore $\frac{\omega^2 M^2}{r_a} + \frac{\omega^2 L M \mu}{r_a}$. It must be noticed that this result is in no way dependent upon the tuning of the aerial, although of course if the aerial circuit is in resonance with the transmitter, its impedance becomes purely resistive and may then be denoted by R , in which circumstances

$$i = \frac{e}{R + \frac{\omega^2 M^2}{r_a} + \frac{\omega^2 L M \mu}{r_a}}$$

and obviously will be greater than $\frac{e}{R}$, the value in the absence of reaction, if M is negative in

sign and does not exceed the value μL . The term $\frac{\omega^2 L M \mu}{r_a}$ is then inherently negative in sign and is the negative resistance introduced into the circuit by the regenerative action. This decrease of total resistance results in an increase in the magnification of the circuit, and consequently the aerial circuit is more selective when used with reaction than in the absence of the latter. To illustrate this, consider the aerial circuit to be tuned to a frequency $\frac{\omega}{2\pi}$ where $\omega = 5 \times 10^6$, e.g. $L = 160 \mu H$, $C = .00025 \mu F$, and $R = 50$ ohms. The resonance curve for this circuit is given

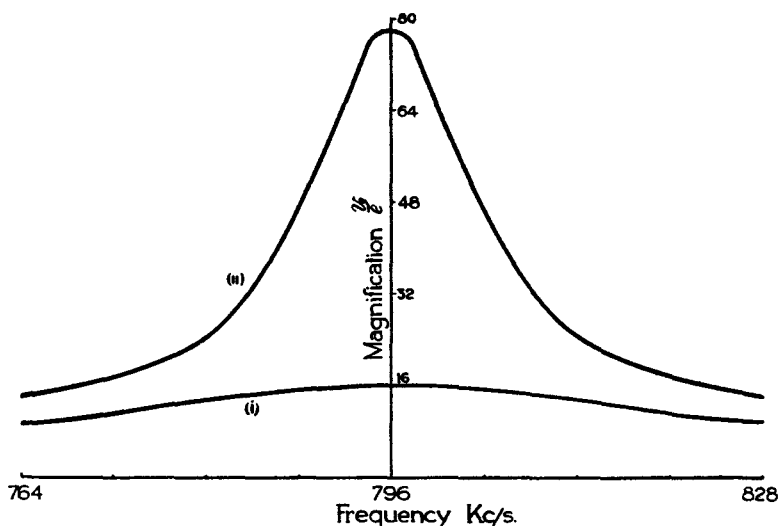


FIG. 22, CHAP. X.—Resonance curve of aerial circuit, with and without reaction.

CHAPTER X.—PARAS. 31-32

in fig. 22 curve (i). Now if coupled to a triode of $r_a = 10,000$ ohms, $\mu = 10$, by a negative mutual inductance of $10 \mu H$ between the aerial and reaction coils, the additional resistance added to the aerial will be

$$\frac{\omega^2 M^2}{r_a} = \frac{25 \times 10^{12} \times 100 \times 10^{-12}}{10^4} = .25 \text{ ohm.}$$

While the amount by which the resistance is reduced, i.e. the negative resistance, is

$$\begin{aligned} \frac{\omega^2 M L \mu}{r_a} &= \frac{25 \times 10^{12} \times 10 \times 10^{-6} \times 160 \times 10^{-6} \times 10}{10^4} \\ &= 40 \text{ ohms.} \end{aligned}$$

Hence the effective resistance of the aerial circuit is reduced from 50 to $10 \cdot 25$ ohms, and the resonance curve under these conditions is shown in curve (ii) of fig. 22.

31. The effect of such a reduction in aerial resistance is two-fold. In the first place, as the circuit magnification is increased, the grid-filament voltage due to the signal will be much greater than in the absence of reaction, and in the second, the persistency of the train of oscillations comprising each group of waves will be greater. These effects have already been illustrated in fig. 21. There is a limit to the amount of regenerative amplification thus obtainable, for when the effective resistance of the aerial is decreased below a certain amount the circuit becomes unstable in the following manner. As the reaction coupling is increased the dynamic load in the anode circuit changes, and this causes a shift in the operating point on the $I_a - V_g$ curve, which connotes some slight change in the value of r_a . An alteration in the effective damping of the input circuit follows, giving rise to a further change in the operating point, and so on. The ultimate result is that the anode A.C. resistance, r_a , and oscillatory anode current i_a tend to assume values at which the damping of the input circuit is exactly neutralised and the circuit becomes a form of oscillation generator. Some such process invariably takes place if an attempt is made to reduce the damping nearly but not quite to zero by critical adjustment of the reaction coupling.

The auto-heterodyne

32. The fact that oscillations can be maintained in this circuit without detriment to its rectifying properties suggests a means of dispensing with the use of a separate heterodyne for C.W. reception. In practice, the circuit given in fig. 20 is frequently employed for this purpose, by increasing the reaction coupling so that oscillations occur. As it is necessary for the frequency of the local oscillation to differ from the signal frequency by about 500 to 2,000 cycles per second, the aerial circuit must be detuned by this amount, for the frequency of the generated oscillation is practically equal to the natural frequency of the aerial circuit. This results in a slight reduction of aerial current due to the distant transmitter and of induced E.M.F. in the inductance L , consequently the oscillatory grid-filament voltage is slightly reduced, but at the frequencies employed in modern communication this effect is negligible. At the very low radio frequencies formerly employed for long distance communication however, the effect was of such magnitude that the autodyne receiver could not be usefully employed. As an example, consider the reception of a transmitter operating on 30 k.c/s., the telephone note to be 1,000 cycles per second. The autodyne receiver would then be tuned to $30,000 \pm 1,000$ cycles per second, and would be 3.3 per cent. out of resonance. If, however, the transmitter frequency is 1,000 k.c/s., the local oscillator is required to generate $1,000,000 \pm 1,000$ cycles per second and is detuned only 0.1 per cent. Autodyne reception is therefore mainly used for frequencies higher than about 300 k.c/s. When receiving C.W. signals by the autodyne method it is found that for maximum telephone response the reaction coupling must be adjusted to a critical value somewhat closer than that required to maintain the circuit in oscillation, while for I.C.W. or T.T. reception the point of optimum telephone response is just off the point of oscillation.

33. As in any other form of oscillator, oscillations are produced when the energy transfer is just sufficient to reduce the effective resistance of the oscillatory circuit to zero, i.e. when

$$\frac{\omega^2}{r_a} M^2 + \frac{\omega^2 L \mu}{r_a} M + R = 0.$$

Solving this as a quadratic equation to find the critical value of M ,

$$M = -\frac{\mu L}{2} \pm \sqrt{\left(\frac{\mu L}{2}\right)^2 - \frac{R}{\omega^2}}$$

Taking the same constants as in the previous example, i.e. $L = 160 \mu H$, $\omega = 5 \times 10^6$, $R = 50$ ohms, $r_a = 10,000$ ohms,

$$\begin{aligned} M &= -\frac{160 \times 10}{2 \times 10^6} \pm \sqrt{\left(\frac{160 \times 10}{2 \times 10^6}\right)^2 - \frac{50 \times 10^4}{25 \times 10^{12}}} \\ &= -\frac{8}{10^4} \pm \sqrt{\frac{64}{10^8} - \frac{2}{10^8}} \quad (\text{henries}) \\ &= (-800 \pm 786) \mu H \\ &= -1586 \text{ or } -14 \mu H. \end{aligned}$$

Thus any value of negative mutual inductance between 14 and 1586 μH will maintain the circuit in oscillation. If M exceeds the larger value the positive resistance $R + \frac{\omega^2 M^2}{r_a}$ becomes larger

than the negative resistance $\frac{\omega^2 L \mu}{r_a} M$ and no oscillations are produced, while if the connections to the reaction coil are reversed so that the sign of M is positive, both these terms contribute to the damping of the input circuit. It may here be noted that it is usual to speak of "positive reaction" as that which reduces damping, so that positive reaction is obtained with negative mutual inductance and vice versa.

Comparison of separate heterodyne and autodyne methods

34. (i) In comparison with the receiver using a separate heterodyne, the autodyne receiver has the important advantage that only one tuning adjustment is necessary, instead of two, but this is to some extent off-set by the necessity for keeping the reaction adjustment near the critical position as the tuning adjustment is varied. Nevertheless, searching for a signal on a frequency for which the aerial adjustments are only very approximately known is more easily carried out with an autodyne than with a local oscillator, unless the latter is accurately calibrated. The most sensitive type of C.W. receiver, which however requires very skilful handling, is one in which a local oscillator is used to produce the heterodyne beat, while the detector itself is operated with regenerative amplification. This will be best appreciated after some experience of C.W. reception, e.g. with a receiver of the type described in paragraphs 41-43, using a syntoniser (Stores Ref. 10A/3040) or wavemeter W.39 (Stores Ref. 10A/7156) as a local oscillator. The comparative simplicity of control of the autodyne receiver has in the past rendered it particularly suitable for use in aircraft, but it is now chiefly used where portability is a prime factor, e.g. in pack sets.

(ii) A further disadvantage of the autodyne system (when used in the manner outlined above) is the fact that the aerial circuit must radiate some energy at the frequency of the oscillation maintained by the valve. This is detrimental for two reasons, first it may cause interference in neighbouring receivers, and second, it may divulge the presence of the radiating receiver to an enemy. For these reasons, the modern form of autodyne receiver generally incorporates a radio-frequency amplifying stage, which also serves to prevent this radiation or at any rate to reduce it to a negligible amount. Such receivers are dealt with in Chapter XI.

CHAPTER X.—PARA. 35

(iii) While it is impossible to receive C.W. signals without either a mechanical interrupter or a local oscillator at the receiving station, I.C.W. signals and indeed any form of modulated wave may be received upon an oscillating receiver designed for C.W. reception. The effect of the local oscillation is to cause the total input voltage to the detector to be greater than in its absence, and the efficiency of rectification is increased, so that a stronger signal usually results. The characteristic note of the I.C.W. signal becomes distorted to such an extent that all indication of its original quality is eliminated, and the only advantage of I.C.W. over C.W. communication is lost. For this reason, as well as to avoid unnecessary interference with other stations by radiation from the receiving aerial, a receiver should never be used in this condition for the reception of other than C.W. signals.

Anode bend rectification

35. (i) Instead of making use of the curvature of the $I_g - V_g$ characteristic, the curvature of the $I_a - V_g$ characteristic may be utilised in order to achieve the desired rectification of a signal E.M.F. The $I_a - V_g$ characteristic of a tungsten-filament triode has two regions of

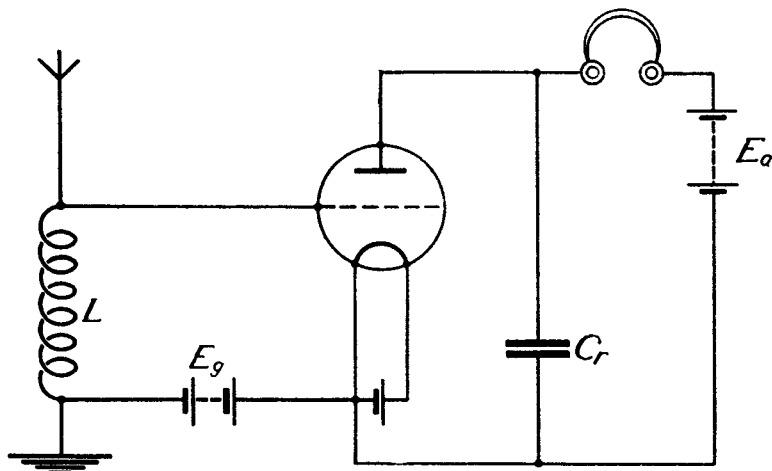


FIG. 23, CHAP. X.—Lower anode bend rectification.

well-marked curvature corresponding to low values of anode current and saturation current respectively; rectification will take place if the mean operating point is located in either region. Valves having oxide-coated filaments or indirectly-heated cathodes do not exhibit a marked saturation current, and as such valves are in almost universal use, it is proposed to consider first the form of rectification known as lower anode bend rectification. A suitable receiving circuit is shown in fig. 23 which may be compared with the circuit used for cumulative grid rectification. It differs from the latter in that the grid condenser and resistance leak are omitted, and a bias battery is provided, by which the normal potential of the grid may be made negative with respect to the filament. The value of this negative potential should be such that in the absence of a signal E.M.F. the anode current is maintained at a value near the point of greatest curvature of the $I_a - V_g$ characteristic. A typical $I_a - V_g$ curve is shown in fig. 24. The mean operating point P is maintained by the grid bias battery, which has an E.M.F. of 6 volts, and as no grid current flows, the grid will be maintained at a negative potential of 6 volts with respect to the negative end of the filament. In the absence of a signal the anode current is about 2 milliamperes in the particular conditions illustrated.

(ii) Let us now consider the reception of a single group of an I.C.W. transmission, which sets the receiving aerial into oscillation, and will develop a corresponding oscillatory voltage

in the aerial tuning inductance L , the peak P.D. being say 3 volts. On each positive half-cycle of grid oscillatory voltage, the anode current will increase to 2 milliamperes, and will fall to zero when the grid filament potential reaches some negative value, (about $7\frac{1}{2}$ volts in fig. 24). The anode current will therefore vary at the frequency of the signal voltage, which is too high to cause vibration of the telephone diaphragm directly. The anode circuit of the valve consists of the H.T. battery, the telephones and the anode-filament path of the valve itself, the telephones and battery being shunted by the condenser C_r which is called the telephone condenser. This condenser fulfils exactly the same purpose as the reservoir condenser in diode rectification. Under pre-signal conditions, the P.D. between its plates is that of the H.T. battery, viz. 100 volts, for the battery charges it to this voltage as soon as the circuit is completed. The upper plate of the condenser is therefore 100 volts positive with respect to the lower.

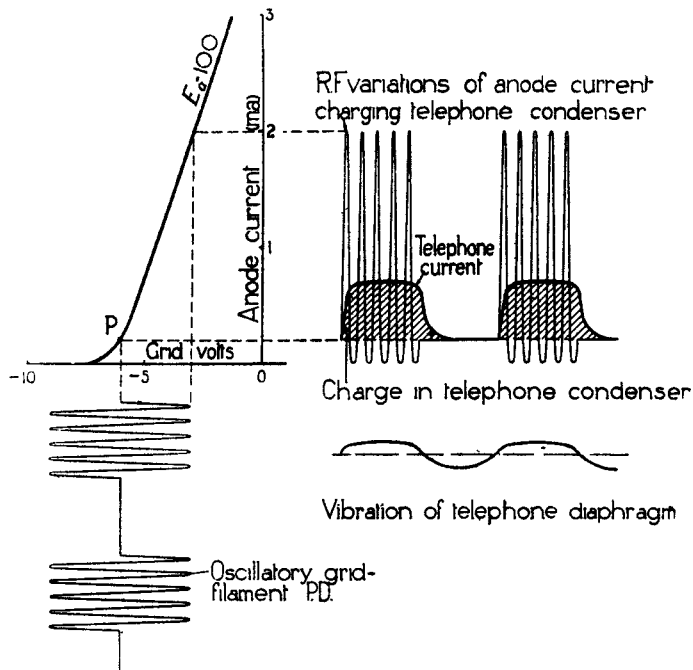


FIG. 24, CHAP. X.—Action of lower anode bend rectifier.

(iii) The opposition of the telephone windings to a steady current is merely the D.C. resistance, but to an oscillatory current the impedance offered is extremely high. During the positive half-cycles of oscillatory grid-filament P.D., the resulting increase of anode current will not flow through the windings, but will be forced to pass into the telephone condenser C_r . As an increase of anode current is equivalent to an increase in the rate at which electrons arrive at the anode, this implies an electron flow into C_r at its upper plate, and the P.D. between the plates is reduced. During negative half-cycles of grid-filament voltage the resulting decrease of anode current is less than the increase on the positive half-cycle, so that during a group of waves the P.D. across the telephone condenser C_r tends to fall below that of the H.T. battery. This cannot occur because the excess electrons in the upper plate of C_r flow steadily to the positive pole of the battery during the whole period, so that for each group of waves a unidirectional current is established in the telephone windings, causing a single inflection of the diaphragm, hence the latter is set in vibration at a frequency corresponding to the number of wave-groups per second. The variations of grid potential, anode current, and telephone current are shown in fig. 24.

CHAPTER X.—PARAS. 36-37

36. If a reaction coil is included in the anode circuit, lower anode bend rectification may be combined with regenerative amplification as already described, while if the reaction coupling is sufficiently increased, autodyne reception of C.W. waves is possible. Owing to the fact that the valve is operated in a region where the anode A.C. resistance is higher than normal, the reaction coupling must be greater than in a similar circuit arranged for cumulative grid rectification. The reaction is also more difficult to control, because the degree of coupling necessary to cause oscillations to commence is much greater than that required to maintain them once they have started. This effect is often referred to as "overlap" or "back-lash" in the reaction control (paragraph 43). It is undesirable because it causes difficulty in ensuring that the receiver is adjusted for optimum conditions, i.e. just over the point of oscillation for autodyne reception or just "off" the point of oscillation for regenerative amplification. The sensitivity of anode bend rectification is generally only about one-half that of a similar receiver using cumulative grid rectification but, provided that the negative grid bias is sufficient to prevent the flow of grid current, and so avoid the damping it imposes, the selectivity is generally somewhat better.

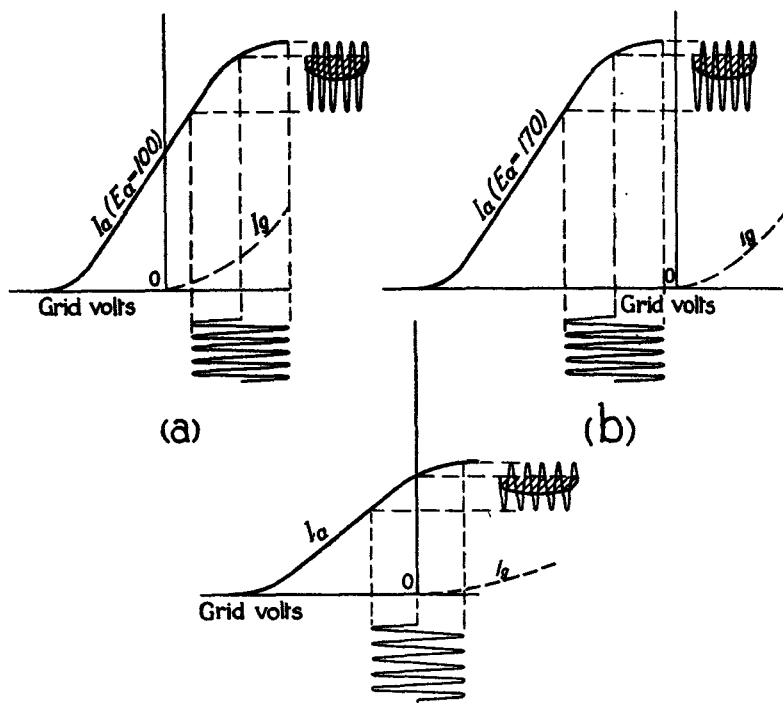


FIG. 25, CHAP. X.—Operating conditions with various types of upper anode bend rectifier.

Upper anode bend rectification

37. Little need be said about upper anode bend rectification. Its mode of operation follows from the previous discussion, observing however that instead of selecting an operating point near the foot of the characteristic by adjustment of grid bias, a point near the saturation bend is selected by one of several methods.

(i) Normal H.T. voltage being used, a positive bias may be applied to the grid, giving the operating conditions of fig. 25a. The disadvantage of this method is the heavy grid current which flows, and the resulting damping thrown upon the input circuit.

(ii) Using negative grid bias, an extra large H.T. voltage is applied, shifting the saturation bend of the curve to the left as shown in fig. 25b. The high value of H.T. voltage required for this method renders the system uneconomical.

In both the above schemes, normal emission is required from the filament.

(iii) Using normal H.T. voltage and zero grid bias, the filament emission is reduced by cutting down the filament current (fig. 25c). The drawback of this method is the reduced slope of the characteristic curve, which causes the rectifier to be even less sensitive than when operating on the lower anode bend.

Owing to the disadvantages of these methods they have fallen into practically complete disuse.

Comparison of cumulative grid and anode bend rectification

38. The relative merits of cumulative grid and anode bend rectification, so far as telegraphic reception is concerned, are briefly as follows. Bearing in mind that in the former type, rectification takes place in the grid circuit, the location of the mean anode current is on the straight and steepest portion of the $I_a - V_g$ curve, and the audio-frequency impulses set up in the grid circuit are magnified by the amplifying properties of the valve. The conditions in fact correspond to rectification followed by audio-frequency amplification, the latter process being efficiently performed owing to the operation taking place at a suitable point on the $I_a - V_g$ characteristic. The anode circuit impedance (i.e. the telephones in a simple receiver without A.F. amplification) is also easily matched to the internal impedance by the use of so-called high-resistance (really high inductance) telephones. In the anode bend rectifier, on the other hand, grid current need not occur, and the valve operates essentially as a radio-frequency amplifier having an unsymmetrical output voltage, which implies that the waveform contains a rectified component. As the location of the mean operating point on the $I_a - V_g$ characteristic is near one of the bends where the slope is comparatively small, the conditions are not favourable for efficient amplification, while the large value of the anode A.C. resistance in this region necessitates a very high impedance load in the anode circuit for correct matching, ordinary high-resistance telephones not being suitable. As a result, for small input voltages, say below 0.5 volts (peak value) the cumulative grid rectifier will give a much louder telephonic response than the anode bend rectifier, but as the input voltage is increased, this advantage becomes of less importance. In most practical receivers cumulative grid rectification is employed.

Simultaneous cumulative grid and anode bend rectification

39. In certain circumstances it is possible for both grid and anode rectification to occur simultaneously, but this is rarely an advantage. If in a cumulative grid rectifier the mean operating point is situated at a point of curvature of the $I_a - V_g$ characteristic as shown in fig. 26a, bottom anode bend rectification must occur to some extent. Now the effect of grid

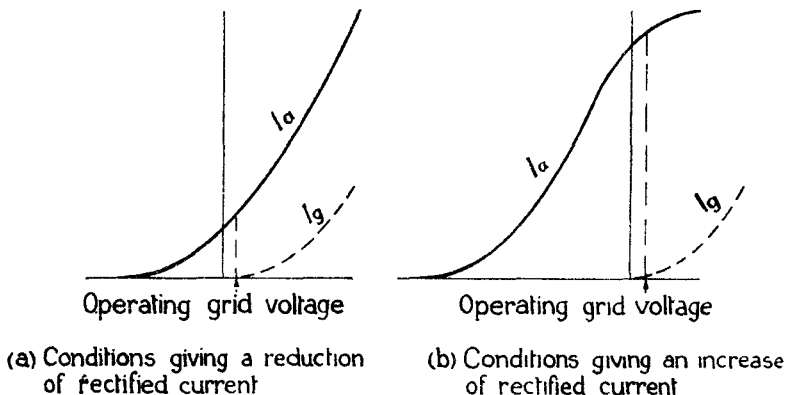


FIG. 26, CHAP. X.—Simultaneous cumulative grid and anode bend rectification.

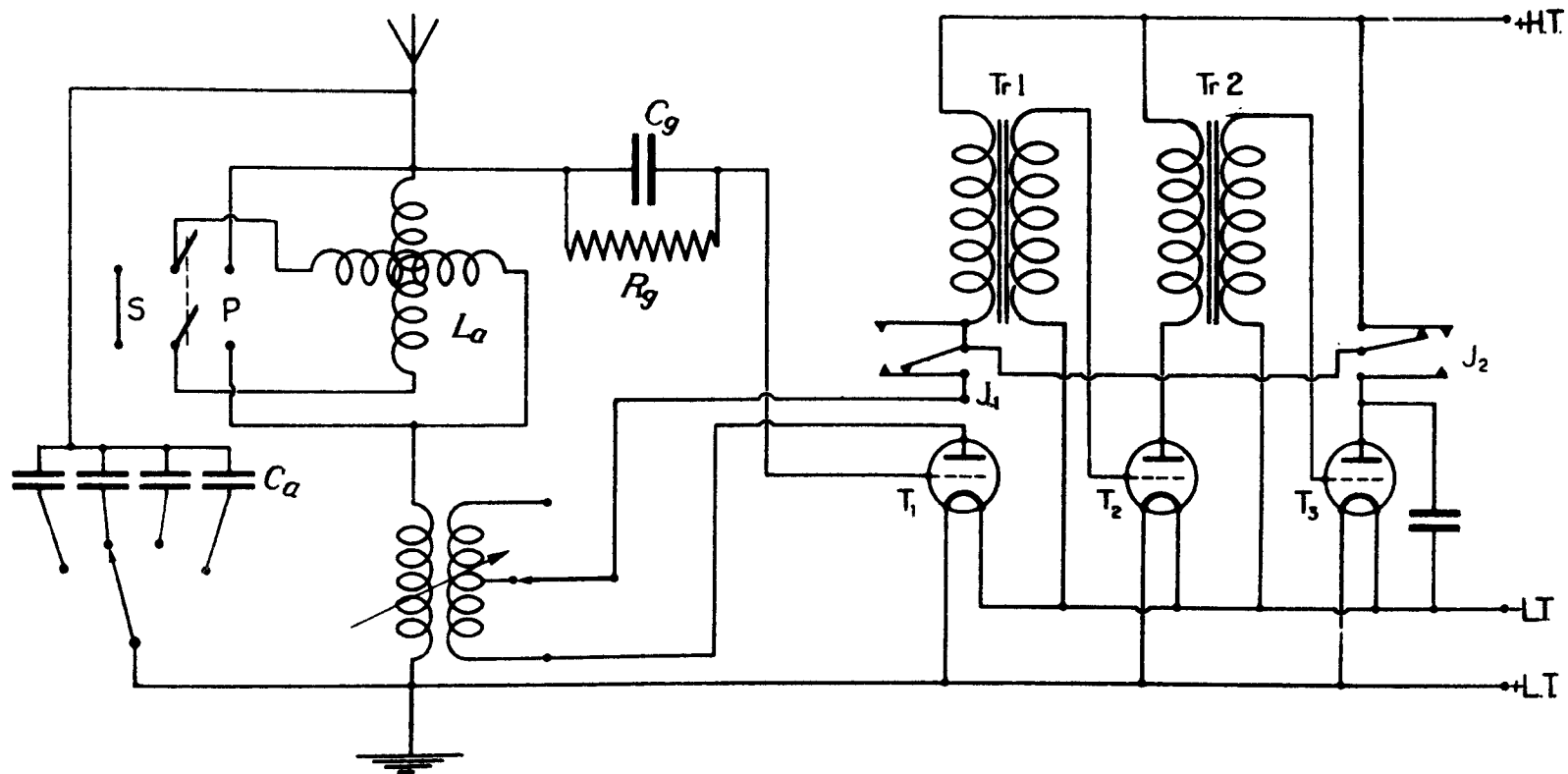
CHAPTER X.—PARAS. 40-41

rectification is a fall of anode current for each group of waves, whereas the effect of bottom anode bend rectification is to give an increase of anode current for each group and the resultant change of anode current will be equal to the difference between the changes due to the two separate causes. It will therefore be less than in a properly adjusted cumulative grid rectifier. This phenomenon is likely to occur if a run down H.T. battery is used. If however a cumulative grid rectifier is so adjusted that the mean operating point on the $I_a - V_g$ characteristic is near the region of saturation current, fig. 26b, the change of anode current due to rectification in the grid circuit will be a decrease, and a further decrease will occur as a result of upper anode bend rectification. The two effects will thus combine to give a greater rectified current than would be given under normal operating conditions. This condition rarely arises in practice.

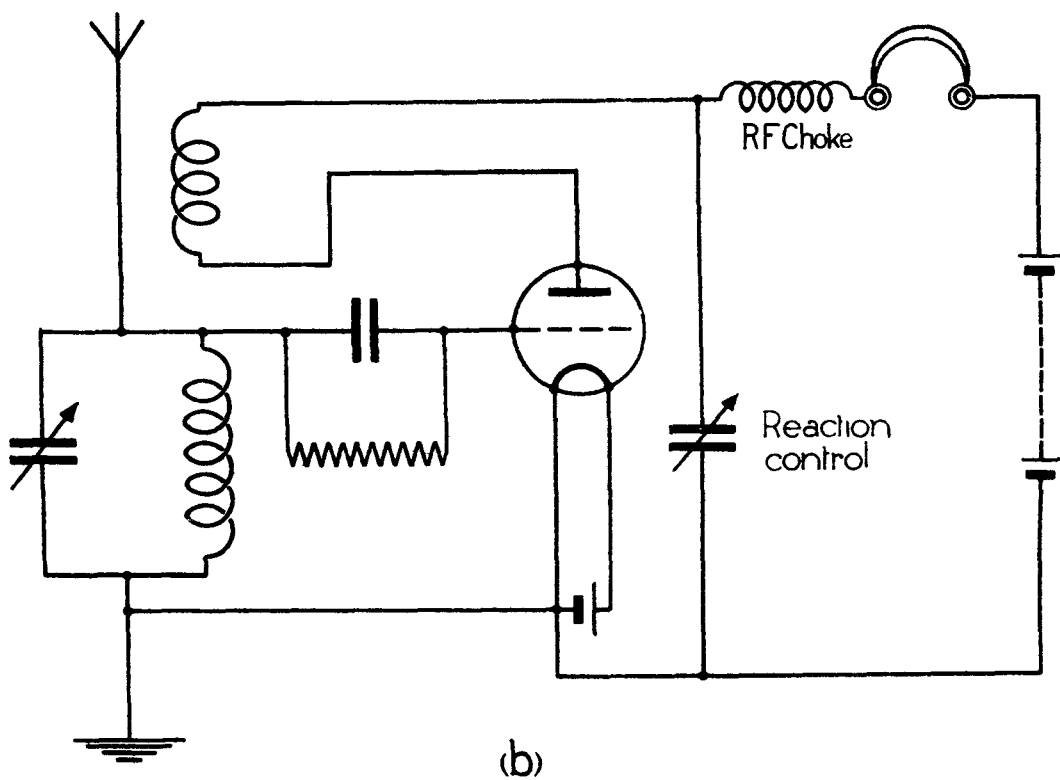
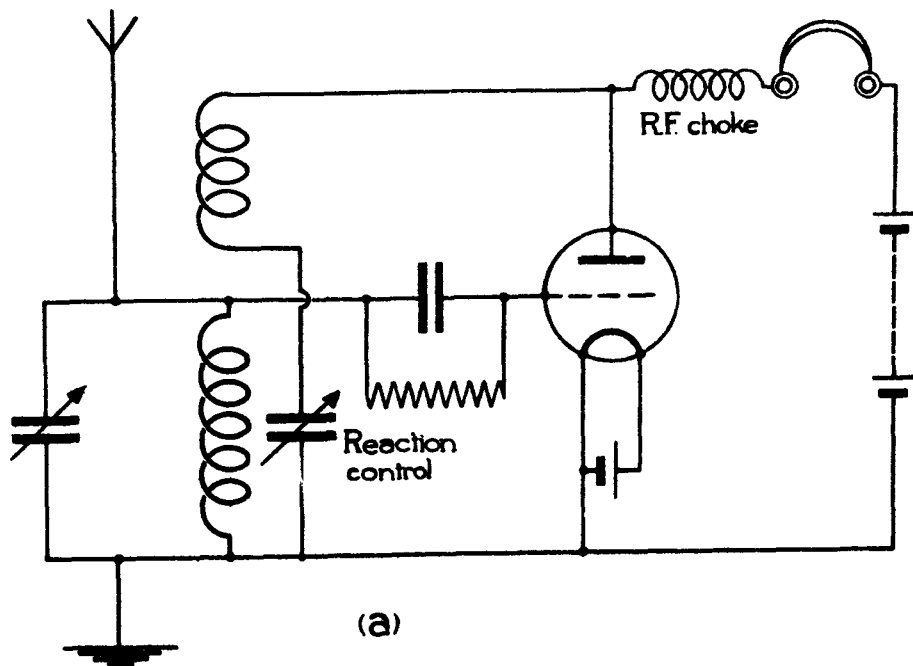
SIMPLE W/T RECEIVERS

40. The circuit diagram of an early aircraft receiver which is capable of either C.W. reception by the autodyne method or regenerative amplification of I.C.W. signals, is shown in fig. 27. Cumulative grid rectification is employed, while aerial tuning is accomplished by means of the variometer L_a , the windings of which may be placed either in series or parallel by means of a switch. When the windings are in series, the frequency range is from 300 to 120 kc/s, while with the windings in parallel the frequency range is from 600 to 200 kc/s. The whole of the aerial tuning inductance is wound with stranded cable in order to keep the radio-frequency resistance as low as possible. The capacitance of the average aerial used with this receiver is about $0.0002\mu F$, and the inductance alone is insufficient to provide the required frequency range, hence a bank of condensers is provided any one of which may be placed in parallel with the inductance. The arrangements for regenerative amplification, etc., are as follows. In series with the aerial tuning variometer is a coil consisting of a few turns of stranded cable which also forms a portion of the aerial circuit. This coil is wound on the stator of a pair of concentric formers, one of which is capable of rotation inside the other. This device is similar in appearance to a variometer, except that the two windings are not directly connected in any way, that carried by the rotor being the reaction coil and being connected in series between the anode of the detector valve and the telephone jack J_1 . Variation of reaction coupling is performed by rotation of the inner coil with respect of the outer an external knob and scale graduated from 0° to 180° being fitted. Since the rotor is capable of 180° rotation, the true zero inductive reaction is at 90° , when the coils are perpendicular to each other, and for any angular setting less than this, the inductive reaction is negative, tending to increase the damping of the aerial rather than the reverse. Owing to considerable capacitive coupling between the two coils, however, zero reaction, including that due to both capacitance and inductance, is generally found at about 45° on the scale. The range switch is arranged to cut out a portion of the anode reaction winding when reception of the higher frequency band is desired.

41. By means of the telephone jacks J_1 and J_2 the telephones may be inserted in the anode circuit of either the first or the third valve of the receiver. The valves T_2 and T_3 are used as audio-frequency amplifiers. When the telephones are plugged into the jack J_2 , the primary winding of the transformer $T_r 1$ is connected in series with the anode circuit of the first valve, just as the telephones are connected in the circuits hitherto considered. The audio-frequency variation of anode current will now set up a varying flux in the iron core of the transformer $T_r 2$, and an E.M.F. will be induced in the secondary winding. This winding is connected to the grid and filament of the second valve, and consequently the secondary E.M.F. of the transformer will cause variations of its anode current at the audio-frequency. A similar action occurs in the transformer $T_r 2$ and triode T_3 , so that the variations of anode current in the latter are a copy of those in the anode circuit of the detector valve T_1 , but are magnified some 50 to 100 times, thus giving a much louder signal in the telephone receivers. The valves T_2 and T_3 are called audio-frequency amplifying valves, and a more complete account of their action is found in chapter XI.



SIMPLE RECEIVER FOR ICW OR CW WAVES



METHODS OF CONTROLLING REACTION

Manipulation of C.W. receiver

42. When a receiver of this kind is used on the ground, a syntoniser is usually also available. This instrument is only calibrated approximately and its scale is marked in wavelength. Its primary use is as a separate heterodyne, and the method of employment is as follows. Tune the receiver to the exact frequency of the desired transmission, using the wavemeter W.39 if possible for this purpose. This instrument is a calibrated oscillator capable of emitting either tonic train (i.e. modulated) or pure C.W., a switch being provided in order to change the type of emission. The wavemeter should be set up to emit tonic train waves in accordance with the instructions on the lid, and the receiver tuned to this emission, reaction being so adjusted that the receiver is near to the point of oscillation but not actually oscillating. The wavemeter should then be set to emit C.W. of the same frequency. This will produce no sound in the receiver telephones, because no beat effect has yet been produced. On setting up and rotating the tuning condenser of the syntoniser, until its frequency is near that of the wavemeter, heterodyne beats will be set up and detected by the receiver; the frequency of the syntoniser can be then brought within 200 cycles or so of the wavemeter frequency by adjusting it to the middle of the dead space. Note the adjustments of both syntoniser and receiver, then switch off the wavemeter and return it to its stowage place. Subsequent reception is performed on the receiver, using the syntoniser to provide the local oscillation. It will be found that the signal strength obtained in this manner is much greater than when the same receiver is used for autodyne reception, while the selectivity is also superior. For preliminary adjustment the syntoniser should be placed as close as possible to the tuning coils of the receiver, although a reduction of coupling may afterwards be found advantageous, and of course slight "trimming" adjustments must be made when the desired station is first heard.

Control of reaction

43. In the simple receiving circuits hitherto discussed, autodyne reception of C.W. or regenerative amplification of I.C.W. is achieved by the use of a reaction coil, and variation of the degree of coupling between anode and grid circuits is performed by varying the mutual inductance of the grid and reaction coils. This is not the best method of reaction control, because a very small change in the relative position of the coils may in certain conditions result in a large change of mutual inductance. In particular it is desirable that the circuit should possess the following property. Starting with no appreciable transfer of energy from anode to grid circuit, that is, negligible reaction, let the coupling be gradually increased until the grid circuit just commences to oscillate. If then, the slightest movement of the reaction control device in the contrary direction causes the oscillation to cease, the circuit is said to have no overlap, while if oscillations, once started, persist in spite of an appreciable reduction of reaction coupling the circuit is said to possess overlap. The condition of no overlap is preferable to the condition of overlap, because for I.C.W. reception the receiver can be brought very near to the point of oscillation without instability, the latter term being in general use to denote a tendency to break into oscillation when some slight irregularity occurs in the operating conditions. e.g., a movement of the operator which may alter the capacitance of some portion of the receiver with respect to earth. Again when receiving C.W. signals by the autodyne method, it is usually found that maximum sensitivity is achieved when the amplitude of the local oscillatory current is a maximum; this condition is generally obtained with the weakest reaction coupling which is capable of maintaining the oscillation. A receiver having overlap can therefore never be operated at the point of maximum sensitivity either as a C.W. or I.C.W. receiver.

44. The usual method of reducing this tendency is to use a fixed amount of mutual inductance between grid and anode circuits, and to control the reaction by controlling the amplitude of the radio-frequency component of anode current. A typical arrangement is shown in fig. 28a. Here the aerial and input circuits are connected to the valve in the usual manner, cumulative grid rectification being employed. A radio-frequency choke is inserted in series with the telephones so that no appreciable oscillatory current may flow in this direction. This choke is necessary owing to the self-capacitance of the telephone windings which would otherwise provide a path

CHAPTER X.—PARA. 45

for radio-frequency current. The reaction coil consists of a winding which is carried on the same former as the grid tuning inductance, the mutual inductance between grid and reaction coils depending only on the relative number of turns and on their separation. The reaction coil, in series with a variable condenser, is connected between the anode and filament of the valve, and the condenser is used as a reaction control, the action being as follows. If the reaction control capacitance is set to zero, no radio-frequency variation of anode current will flow in the reaction coil and therefore no transfer of energy from anode to grid circuits will occur, i.e. no regenerative amplification will take place. As the capacitance is increased, however, the impedance of the circuit is reduced and an increasing radio-frequency current will flow through the reaction coil, so that an appreciable E.M.F. will be induced in the grid circuit giving regenerative amplification of I.C.W. signals; still further increase of capacitance will result in additional energy transference and self-oscillation will occur. Overlap is avoided by careful design of the reaction coil and by choice of a suitable maximum capacitance for the reaction control condenser. An alternative scheme of connection is shown in fig. 28b in which the action is precisely the same.

Use of negative reaction

45. When a receiver is intended for employment under specially arduous conditions, for instance, as a portable receiver such as is used in battery ground stations for army co-operation, certain difficulties arise. In the first place, it is preferable to avoid the use of variable condensers of ordinary design, which are of somewhat delicate construction and may easily be damaged by the vibration to which they must be subject during transit on a gun limber over rough ground.

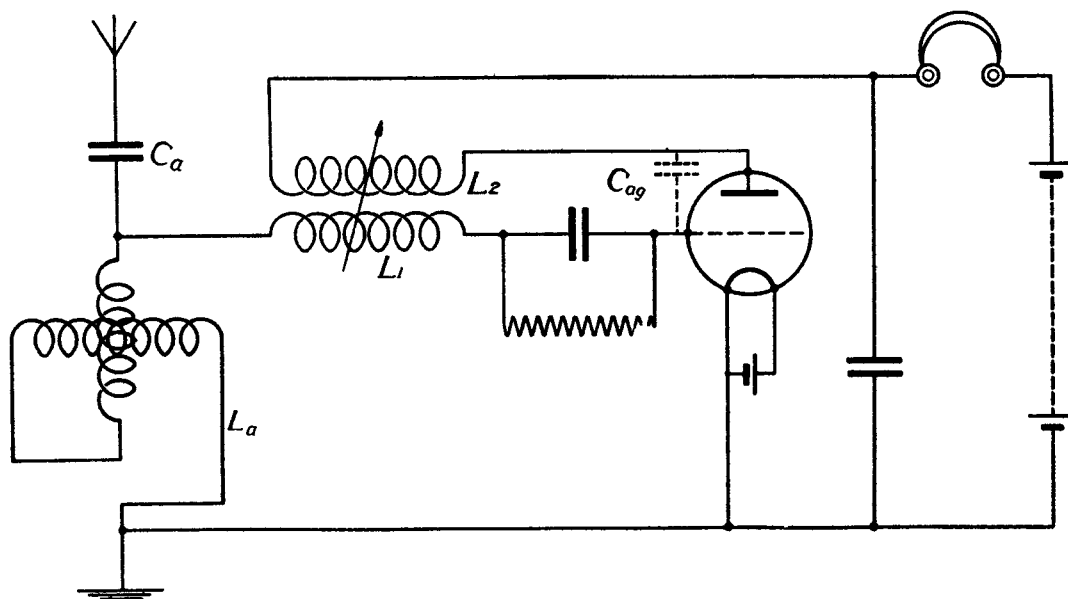


FIG. 29, CHAP. X.—Receiver with combined capacitive and negative inductive reaction.

In the second place, if the variable mutual inductance type of reaction control is employed, the inherent compactness leads to difficulties, for owing to the large and unavoidable capacitance which exists between anode and grid components due to their close proximity, it is found that oscillations may occur even if the reaction and grid coils are separated by the greatest distance possible. On the other hand, if the reaction coil is entirely omitted the sensitivity is poor because the grid circuit is excessively damped owing to the Miller effect. In these circumstances control of the degree of regenerative amplification may be obtained by so connecting the reaction coil

that the effect of increasing the mutual inductance is to impose additional damping upon the input circuit. This type of reaction control was used in an early type of portable receiver for the reception of I.C.W. signals only; the essential portions of the circuit, from the present aspect, are shown in fig. 29. The aerial circuit consists of a fixed condenser C_a and a variometer L_a , the windings of which may be placed either in series or parallel in order to cover a wide frequency range. In series with the grid condenser and leak is a fixed inductance L_1 of very compact design and therefore with little external magnetic field, the inductance being about $1500 \mu H$. The anode circuit is similar to that previously discussed, the reaction coil L_2 being connected in series with the telephones; its electrical and mechanical dimensions are identical with those of the inductance L_1 and it is pivoted upon an arm in such a manner that the coupling between L_1 and L_2 can be varied between wide limits. When the two coils are widely separated oscillations are maintained in the aerial circuit by the stray capacitive coupling existing between grid and anode circuits. As the reaction coil is connected in such a manner that an increase of magnetic coupling increases the damping of the input circuit, an increase of reaction coupling tends to suppress the oscillation and this suppression occurs when the coupling exceeds a certain critical value. The reaction coil can then be locked in this position.

46. The variation of reaction may be shown graphically. In fig. 30 positive reaction is considered to be that which causes a reduction of damping and negative reaction that which increases the damping of the input circuit. A certain amount of positive reaction is present owing to the stray capacitance, this amount being above that necessary to maintain oscillation.

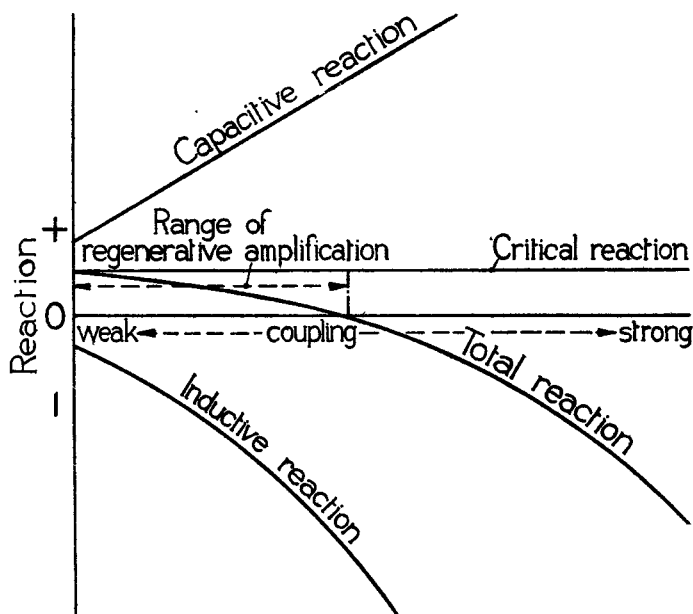


FIG. 30, CHAP. X.—Control of reaction in receiver of fig. 29.

The capacitive coupling between grid and anode circuits increases slightly as the two coils approach each other, so that the total positive (i.e. capacitive) reaction increases as the controlling lever is moved from one end of its travel to the other, as shown in the diagram. The magnetic coupling between the coils, however, is inversely proportional to some power of the distance between the coils; the variation of inductive reaction is therefore somewhat as shown, the total reaction being the algebraic sum of the capacitive and inductive reactions. It will be seen that by suitable adjustment of the circuit constants and operating potentials the reaction

CHAPTER X.—PARAS. 47-48

may be made to approach in a very gradual manner the critical value at which oscillation commences. Intelligent operation of this device is essential if optimum results are to be obtained. In particular, it should be realised that although the valves fitted in the portable receiver referred to are designed for a filament voltage of 2 volts, the inert battery used for the filament supply has an initial E.M.F. of nearly 3 volts, gradually falling during prolonged use, and returning nearly to its original value after a period of rest, unless the battery is near the end of its useful life. It is, therefore, not only unnecessary but a positive disadvantage to move the filament rheostat to the full "out" position when coming into action. The rheostats, and in particular the detector valve rheostat, should be set at a position giving both ample signal strength and a smooth control of reaction, the latter being only obtained when the anode current of the detector valve is not excessive. Only by attention to this detail will satisfactory operation of the reaction control be ensured. The fact that this receiver tends to break into the squegger type of oscillation (Chapter XI) when the reaction control is mishandled, is also an advantage, for it prevents the operator from using the receiver in an oscillating condition in an endeavour to obtain increased signal strength. In this connection it must be noted that operation in this way not only causes the characteristic note of the I.C.W. transmitter to be unrecognisable, so that "zone calls" are indistinguishable from ordinary signals, but renders the battery station liable to location by enemy direction-finding apparatus.

Send-receive switches

47. Where a transmitter and receiver are installed conjointly as is usual in aircraft, it is obviously desirable to utilise the same aerial both for transmission and reception. To a certain extent this usage also applies to ground stations, except where the transmitter is erected at some distance from its receiver and is controlled via land line by the receiving operator. When the same aerial is to be used for both reception and transmission, a send-receive switch is used to change the aerial connection from transmitter to receiver, and it is usual to combine this function with others. In an aircraft transmitter the send-receive switch will usually perform the following duties :—

(i) In the "transmit" position :—

- (a) Connect aerial to transmitter aerial coil.
- (b) Complete positive and negative H.T. supply from generator to transmitter.
- (c) Complete positive L.T. supply to transmitter.
- (d) Disconnect receiver from aerial.
- (e) Transfer telephones from receiving circuit to side tone unit (see Chapter XII) in the case of I.C.W. or R/T transmitters.

(ii) In the "receive" position :—

- (a) Break the connections (a), (b), (c) above.
- (b) Connect aerial to receiver.
- (c) Transfer telephones from side tone unit to anode circuit of receiver.

Listening through

48. In signalling by wireless telegraphy it is often desirable that the receiving operator shall be able to interrupt the transmitting operator at any point during the transmission, either because a portion of the message has been missed, or to give urgent operating instructions. This entails that the receiver in use by the transmitting operator shall be operative at all times except when the transmitting key is actually pressed, a requirement which obviously is not met by the use of a send-receive switch as described above. The original solution of this problem was to provide an electrically operated switch which on pressing the transmitting key automatically performed the functions enumerated under (i) (a), (b) and (d) above, while on releasing the key the operations (ii) (a) and (b) were performed. This device is known as a listening-through key and is still in use to a limited extent. It has the disadvantage that it is difficult to attain a

high operating speed owing to the inertia of the moving parts. A device which entails no mechanical motion whatever has supplanted this in modern service sets. The aerial is permanently connected to the transmitting aerial circuit, and the receiver is capacitance-coupled to the latter circuit by means of a very small condenser, the arrangement being as in fig. 31. It will be noted that during actual transmission the whole of the voltage across the aerial coil is applied to the coupling condenser and receiver aerial circuit in series. Even in the most favourable instances, i.e. on high radio-frequencies, with a comparatively small inductance in the transmitting aerial circuit and with low aerial current, the P.D. set up across the input circuit of the receiver will be much higher than that caused by a normal signal. Unless suitable precautions are taken this excessive voltage will probably damage the receiver coils, condensers and valves, and in order to prevent this a limiting valve is fitted.

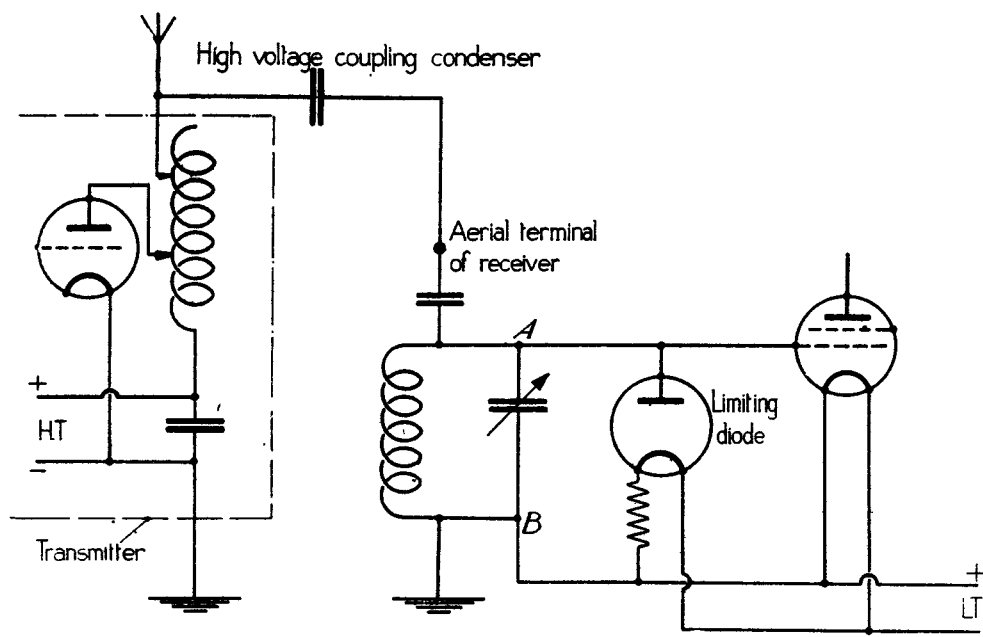


FIG. 31, CHAP. X.—Listening-through device.

49. The limiting valve consists of a diode (or a triode with grid and anode linked together), which is connected in parallel with the tuning condenser in the receiver aerial circuit. The anode is maintained at a small negative potential with regard to the filament, say -1 volt. During ordinary reception, the peak oscillatory voltage across the points *A* and *B*, to which the diode is connected, never reaches one volt, and the anode never becomes positive with respect to the filament; hence anode current will not be established at any point in the cycle, and the receiver operates in exactly the same manner as it would with the diode removed. During transmission, however, the peak value of the induced E.M.F. in the tuned circuit of the receiver may be greatly in excess of one volt, and during the positive half-cycles the anode of the diode becomes highly positive with respect to the filament. An anode current is therefore established, and the anode-cathode space is equivalent to a resistance connected between the points *A* and *B*, the value being comparatively low, e.g. about 400 ohms. The damping imposed on the tuned circuit by this resistance in parallel with the tuning condenser is equivalent to that imposed by a series resistance of thousands of ohms, so that the oscillatory voltage across the points *A* and *B* is never built up to an excessive value. Alternatively, if the diode is considered practically to "short-circuit" the points *A* and *B* when anode current is passing, it is obvious that no P.D.

CHAPTER X.—PARAS. 50-53

can exist between these points and the whole of the voltage developed in the transmitter aerial coil is thrown across the high voltage coupling condenser. It will be observed that the phenomenon is precisely the same as in the case of damping caused by excessive grid current in a receiver, but to a much greater degree.

50. This arrangement has certain advantages and disadvantages over the mechanical listening-through key ; the higher speed of signalling has already been mentioned. The further advantages are :—

(i) Although when the transmitter is operated the voltage applied to the receiver is quite small, it is not zero, and the receiver can be accurately tuned to the transmitter frequency.

(ii) If the receiver is independently tuned to a distant transmitter, the local transmitter can afterwards be tuned to the same frequency.

(iii) The signal strength and selectivity in the "listening through" position are higher than if the receiver is connected directly to the aerial, provided that the transmitter aerial circuit has very low losses. Circuit arrangements usually provide for this method of reception to be used independently of the position of the send-receive switch ; the principal function of the latter is then to break the power supplies to the transmitter.

The disadvantages are :—

(i) Since the receiver and transmitting aerial circuits are coupled together, signals are only received when the transmitter and receiver are tuned to approximately the same frequency. This is of only minor importance because this is the standard practice in the present W/T organisation.

(ii) Steps must be taken to ensure that the noise level due to commutation, etc., is not excessive, otherwise signals cannot be heard during the "space" intervals.

(iii) If the limiting valve becomes soft in use it may become appreciably conductive under the influence of quite small applied voltages, of the order of those set up in the receiver by signals of moderate strength. The valve will then damp the input circuit during reception and cause a reduction in selectivity and in signal strength. This can easily be tested by removing the limiting valve during reception, and the only remedy is replacement by an efficient diode.

INTERFERENCE

51. It is a matter of everyday observation that when a wireless receiver is in operation, a considerable background of noise may be in evidence. Collectively these undesired sounds are referred to as interference, and in turn may be divided into the following classes. (i) Signal interference. (ii) Atmospheric interference. (iii) Electrical interference (also sometimes called "man-made atmospheric" or "man-made static"). (iv) Amplifier noise. The latter will be considered briefly in chapter XI.

Signal interference

52. This term is applied to interference with reception by signals emanating from radio transmitters other than that with which communication is desired. It is only rarely that an interfering signal possesses a frequency identical with that to be received ; when this is the case directional reception is the only remedy. Provided both frequencies are controlled by the use of master oscillators, a judicious use of the heterodyne principle may be very effective in reducing interference on frequencies which differ from the desired signal by only a few hundred cycles per second. Interference from stations giving a high field strength owing to their proximity

to the receiver and their high power, is reduced to a minimum by the use of selective receivers. The following methods are all in common use, separately or in combination :—

- (i) Regenerative amplification in combination with a separate heterodyne for C.W. reception. This gives a high degree of selectivity but requires skilful manipulation.
- (ii) The use of an aerial circuit having a very high ratio of effective inductance to resistance.
- (iii) Reduction of aerial damping by the use of a series aerial condenser of small capacitance. This is a special application of method (ii).
- (iv) The employment of loose coupling between the aerial and the input circuit to the detector valve.
- (v) Employment of tuned radio-frequency amplifiers. These are dealt with in the following chapter. Method (i) has already been described, and methods (ii), (iii) and (iv) will now be discussed.

53. The resistance of an aerial circuit may be divided into three portions. First, the ohmic resistance of the conductor ; this is reduced to a minimum by using wire of suitable gauge, stranded cable being employed if necessary. Second, the resistance of the earth or counterpoise ; the importance of this has been emphasised in connection with transmitting aerials (Chapter VII), but it is often forgotten that a perfect earth may be rendered useless by a high resistance connection to the receiver, and all metallic contacts at terminals, etc., should always be kept clean and bright. Third, the radiation resistance ; in reception, the electro-magnetic field is the cause of an oscillatory current in the aerial, and re-radiation must take place to some extent. This phenomenon must be distinguished from the radiation which takes place due to the use of an oscillating receiver for C.W. reception. Re-radiation must occur whenever an electro-magnetic wave sets up an oscillatory current in a conductor ; steel-framed buildings, large cranes and even trees act in this way. The phenomenon of re-radiation implies the existence of a radiation

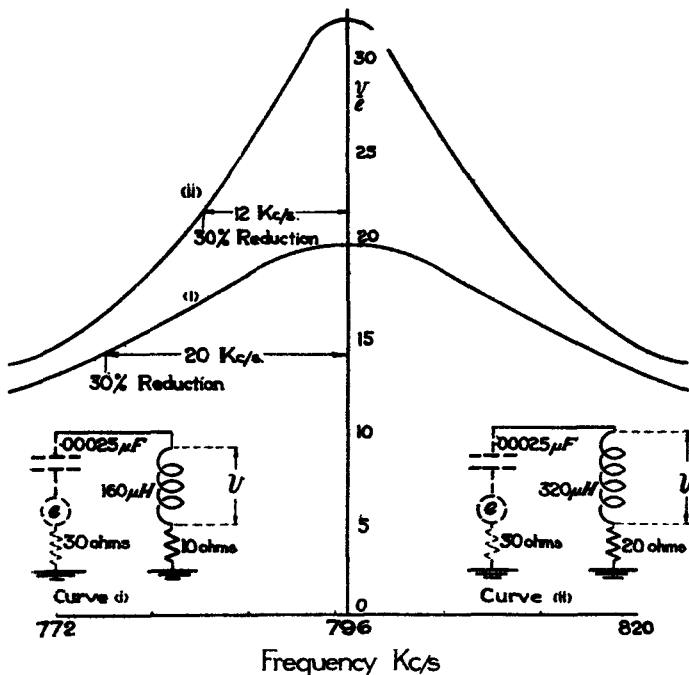


FIG. 32, CHAP. X.—Selectivity ; effect of increasing L/R .

CHAPTER X.—PARA. 54

resistance ; it is usual to assume that the radiation resistance of a given aerial when used for reception is equal to that which it possesses when used for transmission, although this cannot be strictly true because the current distribution is not the same in each case. As an example of the influence of the various components of aerial resistance, the following results of actual measurement may be quoted. An inductance of $160\ \mu H$ and a condenser of $\cdot 00025\ \mu F$ were combined to form a closed oscillatory circuit and its resistance at the resonant frequency was found to be 8 ohms. On removing the condenser and connecting the coil to an aerial and earth of the same effective capacitance, the resistance was found to be 36 ohms. The earth connection in these experiments was not very good, being made on the ordinary water supply system by a metal clamp, but the results are typical of what may be expected when a properly constructed earth system is not used. The calculated radiation resistance of the aerial was only 2 ohms, and it is possible that a large proportion of the 26 ohms resistance which is apparently inherent in the aerial-earth system itself is really due to eddy current and dielectric absorption losses ; these would be reduced by the use of an earth screen as in the case of transmitting aerials. (Chapter VII). The resonance curve of an aerial circuit having the above values of L and C ,

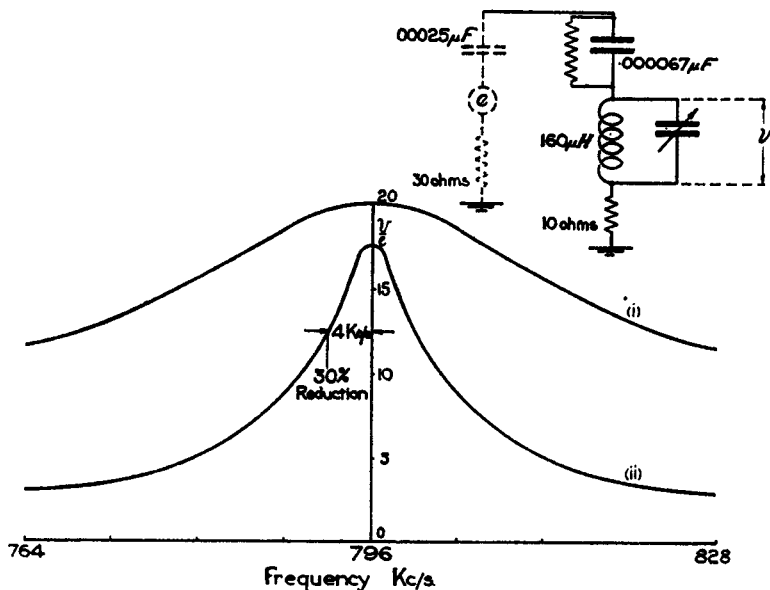


FIG. 33, CHAP. X.—Selectivity ; effect of small series condenser in aerial circuit.

but having a total resistance of 40 ohms, is shown in fig. 32, curve (i). It will be observed that the tuning is very flat, signals 20 k.c/s. off resonance giving an input to the detector equal to 70 per cent. of the desired signal (equal field strengths being assumed).

54. An improvement may be expected by increasing the inductance of the tuning coil to say $320\ \mu H$, inserting a variable condenser in series in order to reduce the total effective capacitance to $\cdot 000125\ \mu H$. This will increase the resistance also, but the ratio $\frac{L}{R}$ will increase in greater proportion. Assuming that the additional inductance increases the resistance to 50 ohms, the resulting resonance curve is shown in curve (ii) of fig. 32. The selectivity is somewhat improved, a 30 per cent. reduction in input voltage being obtained from signals only 12 k.c/s. off resonance. It is not practicable further to improve the selectivity by increasing the value of L without limit, for this entails a corresponding reduction in the series capacitance which eventually approximates to a complete disconnection of the aerial, and the signal strength fails off. Even a

moderate increase of inductance is accompanied by a corresponding increase of distributed capacitance, again limiting the amount of inductance required for resonance. Finally, the larger the coil dimensions become, the more difficult it is to secure controllable reaction and to avoid stray coupling with other inductances in the receiver.

55. The expedient usually adopted in order to attain the equivalent of a high ratio of inductance to resistance is shown in the circuit diagram of fig. 33. A fixed condenser of small capacitance compared with that of the aerial is connected in series, and a variable condenser in parallel with the tuning coil. The effect of the series condenser is to reduce the damping thrown upon the circuit by the aerial resistance. The resonance curves in the diagram are, (i), which is repeated from the previous figures for comparison, (ii), the resonance curve of a circuit having the constants shown in the circuit diagram. A 30 per cent. reduction in input voltage is now obtained from signals only 4 k.c/s. off resonance. This method of attaining a high degree of selectivity possesses considerable advantages, being simple in operation and requiring only the addition of a very small and light component. A possible disadvantage is that the aerial may collect static charges which will gradually raise its potential until it is sufficient to puncture the insulation of the series condenser. This is avoided by connecting in parallel with the latter a high resistance, e.g. 250,000 ohms, which allows the static charge to drain harmlessly to earth.

56. The aerial may be inductively coupled to the input circuit of the receiver, two types of coupling being in common use. In the first type, both the aerial and input circuits are tuned to the desired frequency. A very high degree of selectivity is obtainable by this method, but

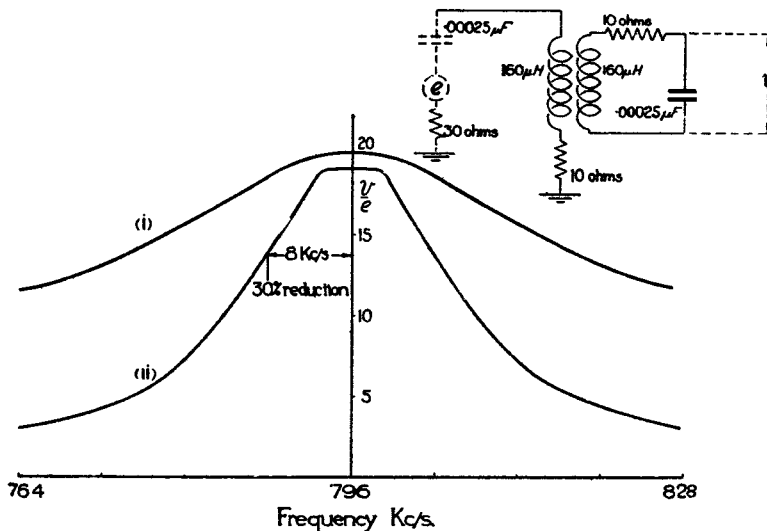


FIG. 34, CHAP. X.—Selectivity ; tuned coupled circuits.

it has the serious disadvantage that if the tuning control of each circuit is entirely independent, some little practice is required in order to achieve exact tuning with the desired degree of selectivity. It is possible to link the two circuits together in such a manner that both circuits are tuned simultaneously by rotation of one control and the circuits are then said to be ganged. Fig. 34, curve (ii) shows the resonance curve of a receiver having an aerial and input circuit of the separately adjustable type, curve (i) being repeated from fig. 32 for comparison. It is not absolutely necessary to tune the aerial circuit, provided that a rather higher percentage of coupling is used ; fig. 35 has been drawn to illustrate this. The aerial circuit has an inductive coupling to the input circuit, but no provision is made for varying either the tuning of the aerial circuit or the degree of coupling. This arrangement is generally referred to as "aperiodic aerial

CHAPTER X.—PARA. 57

coupling." It will be seen from the diagram that the selectivity of the circuit shown is approximately the same as that of the circuit of fig. 34. Actually the centre of the resonance curve is at about 800 k.c/s. instead of 796 k.c/s., i.e., the resonant frequency of the combined circuits is slightly higher than that of the tuned circuit alone. All the resonance curves of fig. 32 to 35 inclusive have been calculated by a graphical method but the degree of accuracy is sufficient for the present purpose, namely, to illustrate the comparative selectivity of certain specific arrangements. The curves show that for C.W. or I.C.W. reception there is very little to choose between the selectivity obtainable by the various means. The requirements of R/T reception are somewhat different and will be dealt with in Chapter XII. Regenerative amplification may be used in conjunction with any of these devices, giving a further increase of selectivity as shown in earlier paragraphs. It appears that from a practical point of view the aperiodic mutual inductive coupling and the use of a small series condenser are equally good; both these methods are in use in different types of service receiver.

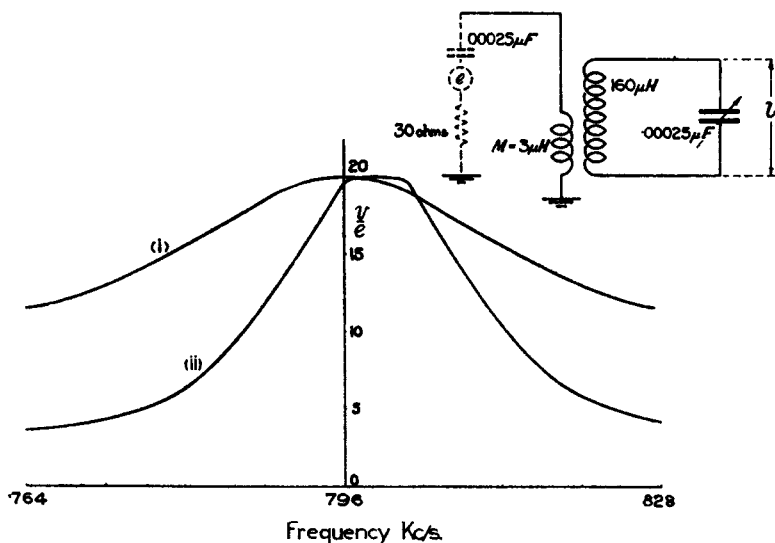


FIG. 35, CHAP. X.—Selectivity; coupled circuits with aperiodic aerial.

The decibel

57. (i) The relation between the strength of a signal at one point in a transmission chain and the strength at any other point depends upon the ratio of the power available at each of the two points. The magnitude of two powers P_1 and P_2 are said to differ by N bels when

$$\frac{P_1}{P_2} = 10^N$$

or

$$N = \log_{10} \frac{P_1}{P_2}$$

For general use a submultiple of this unit—the decibel—is employed, and the gain or loss in the transmission system is n decibels (db) when

$$n = 10 \log_{10} \frac{P_1}{P_2}$$

The two powers to be related need not correspond to dissipation of the same kind of energy, for instance P_1 may be the electrical power delivered to a telephone receiver and P_2 the power

radiated by the diaphragm in the form of sound. One decibel corresponds approximately to the smallest change in sound intensity which can be perceived by the human ear, and the corresponding power ratio is

$$\begin{aligned}\frac{P_1}{P_2} &= \text{antilog } 0.1 \\ &= 1.259,\end{aligned}$$

i.e. an increase in power of 26 per cent. is the smallest which will give an appreciable increase in audibility. The practical value of the decibel arises from its logarithmic nature, for it enables an enormous range of power ratios to be expressed in convenient figures; for example if $\frac{P_1}{P_2} = 10^6$, the gain or loss is 60 db. It also simplifies the calculation of a number of successive gains, for example, if the gain between two points *A* and *B* is n_1 db and between *B* and *C*, n_2 db, then the gain between *A* and *C* is $(n_1 + n_2)$ db. When the two powers under comparison are dissipated in equal resistances the power ratio is proportional to the square of the voltage ratio, or to the square of the current ratio. In these circumstances the gain or loss may be expressed in db by the relation

$$n = 10 \log \left(\frac{E_1}{E_2} \right)^2 = 10 \log \left(\frac{I_1}{I_2} \right)^2.$$

or

$$n = 20 \log \frac{E_1}{E_2} = 20 \log \frac{I_1}{I_2}.$$

This relation must not be used unless the resistance associated with E_1 , I_1 , is the same as that associated with E_2 , I_2 .

(ii) The following examples show how the decibel system is applied :—

(a) In the arbitrary " audibility scale " used in W/T procedure to indicate the strength of signals, an increase of one unit corresponds, very roughly, to an increase of 6 db.

(b) The gain between two points *A* and *B* in a certain transmission system is 30 db. Between *B* and *C* there is a gain of 6 db, while between *C* and *D* there is a gain of .7 db. The total gain between *A* and *D* is $30 + 6 + .7 = 36.7$ db, corresponding to a power ratio of 4,670 to 1. This example shows the advantage of the decibel system. If the above calculation were performed in power ratios, it would be necessary to find the continued product of 1,000, 3.98 and 1.175, instead of the sum of 30, 6 and .7. Again, if between *B* and *C* there is a negative gain, i.e. a loss of 6 db, the overall gain will be $30 - 6 + .7$

or 24.7 db, and the power ratio $1,000 \times \frac{1}{3.98} \times 1.175 = 295$ to 1.

58. In the receiving aerial the initial voltage is dependent only upon the configuration of the aerial and the field strength of the signal, but is independent of the aerial impedance. The output voltage of the complete receiver is that across the telephones, and the latter are not of the same impedance at all audio-frequencies. If we stipulate (i) that the desired and undesired signals are both I.C.W. of the same note-frequency, and (ii) that the response of the receiver (between the input terminals to the first valve and the output terminals) is linear, we may compare the selectivity of the various arrangements shown in fig. 32 to 35 by stating that, for equal field strengths, an interfering signal will be n db below the desired signal. Suppose the latter to be on 796 k.c/s. and the interference on 780 k.c/s. In fig. 32, curve (i) the magnification $\left(\frac{v}{e} \right)$ at 796 k.c/s. is 20 and at 780 k.c/s. is 15.5. The latter is $20 \log_{10} \frac{20}{15.5}$ or about 2 db below the former. In curve (ii) the magnification at 796 k.c/s. is 32 and at 780 k.c/s. is 19; the interference is 4.5 db below the desired signal. In fig. 33 curve (ii) the interference is 11 db, in fig. 34, curve (ii), 8.6 db and in fig 35, curve (ii), 8.3 db, below the level of the desired signal. For the purpose of comparison we may assume that an expert operator may be able to " over-read "

CHAPTER X.—PARAS. 59-60

the desired signal if it is one strength on the audibility scale, i.e. 6 db, above the interference, and it is seen that in the particular circumstances stipulated above, only the last three arrangements are sufficiently selective. In practice, of course, the receiver rarely has a linear response; e.g. a square law rectifier will discriminate in favour of the stronger signal, further improving the selectivity. In C.W. reception, with the same frequency separation, conditions would be much better. In the above circumstances, for instance, to receive the signal of 796 k.c/s. the heterodyne may be adjusted to 795 k.c/s. giving a 1,000 cycle note. The interfering signal would then set up a beat note of 15,000 cycles per second, and to such a high frequency the combination of telephone, ear and brain is comparatively insensitive.

Atmospheric interference

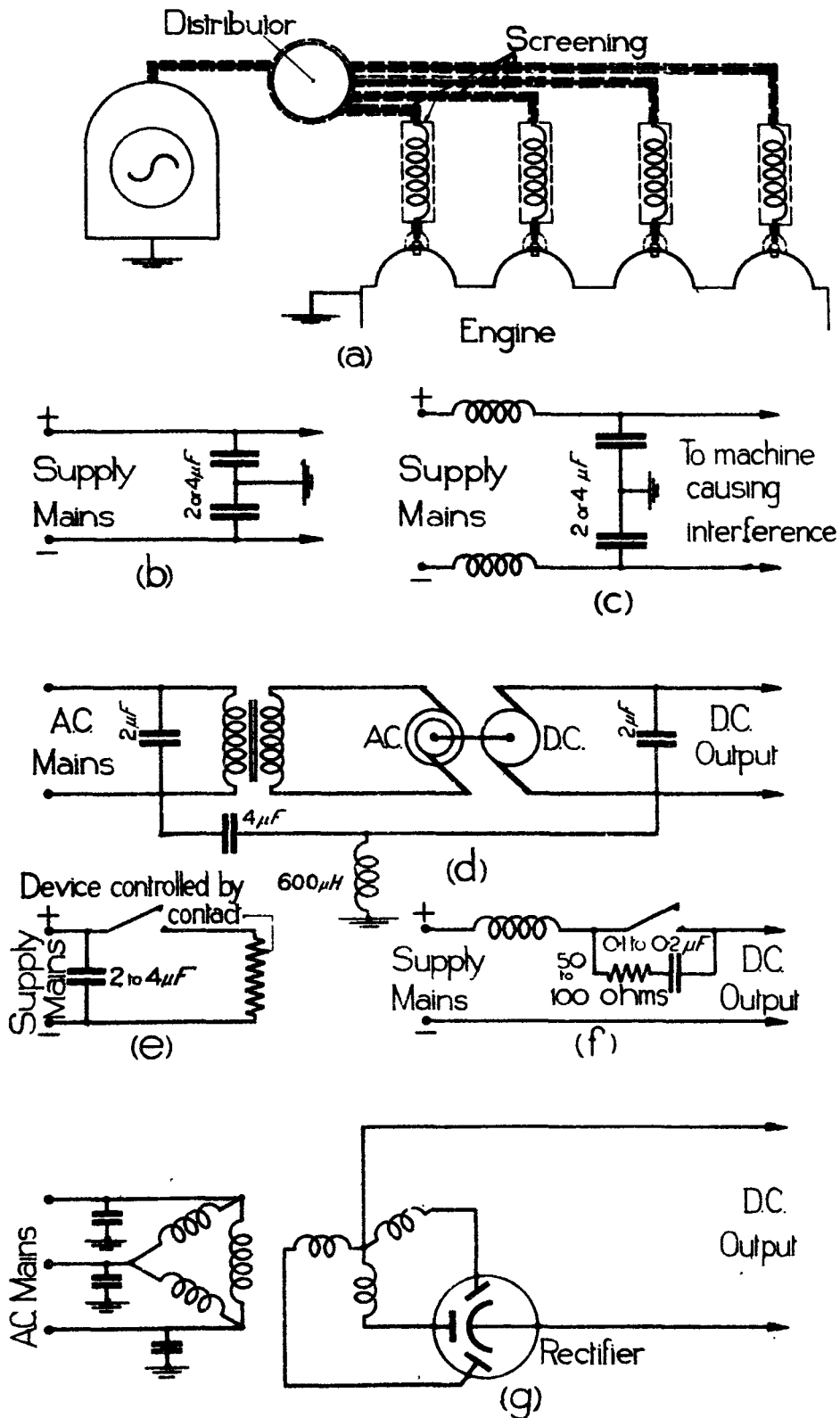
59. (i) This is caused by electro-magnetic waves of natural origin, and its intensity varies greatly with the location of the station—being least in the temperate zones—and with the seasons of the year. Until a few years ago the origin of atmospherics was unknown, but with the aid of the cathode ray oscillograph, the wave-form, field strength and the direction of incidence have been measured in certain instances. As a result of this research it is now believed that most atmospherics originate within the region to which the atmosphere of the earth extends, and are chiefly caused by lightning discharges between clouds, or from clouds to earth, perhaps thousands of miles from the receiving station. The wave form of an atmospheric is almost invariably aperiodic, showing no reversal of polarity, a typical shape being a single pulse rising sharply from zero to a field strength which may be as large as $\cdot 1$ volt per metre, falling to zero in about $\cdot 002$ second. The E.M.F. induced in the aerial circuit then causes it to oscillate at its natural frequency with a decrement depending upon the ratio of inductance to resistance. From another point of view, such an aperiodic pulse may be considered to be built up of a spectrum of oscillations of all frequencies between zero and infinity, the lower frequency components possessing the greater amplitude. The receiver then responds to that portion of the spectrum which is embraced by its resonance curve, and the noise level caused by atmospherics will be directly proportional to the width of the frequency band which the receiver will accept. The position is exactly analogous to the emission from a white-hot solid body of light, which consists of a continuous spectrum in which every colour is present. A colour filter may be transparent to a narrow band of frequencies, e.g. red, and opaque to all others, thus resembling the tuned circuits of a receiver. Such a filter passes most of the light thrown upon it by a red-hot solid, but will pass some light from any incandescent body, and the light which it transmits is red only. In like manner, a wireless receiver always finds in the spectrum of an atmospheric a certain band to which it will easily respond.

(ii) In these circumstances, it appears to be impossible to combat atmospheric interference with W/T signalling by modifications to circuit design, although in the past many futile attempts have been made in this direction, such as the introduction of resistance to damp out disturbances, balancing schemes designed to neutralise the atmospheric without affecting the signal, detuning the receiver, etc. Experience and theory combine to show that such devices are valueless in increasing the ratio of signal strength to noise, and elimination of atmospheric interference must be sought mainly in the following directions :—

- (a) The provision of high signal field-strength.
- (b) Choice of frequency; atmospheric interference being much less troublesome on the frequencies above 3 m.c/s. than below.
- (c) Use of directional receivers and aerial arrays.
- (d) Use of receivers of high selectivity.

Electrical interference

60. Electrical interference may be of three kinds, (i) radio-frequency, ether-borne, (ii) radio-frequency, mains-borne and (iii) audio-frequency. The first two are due to the fact that electric oscillations are established in every circuit of less than critical damping whenever the electrons



SUPPRESSOR DEVICES

therein receive acceleration; the frequency of the oscillation is determined by the circuit constants. In practice it is found that radio-frequency oscillations of a heavily damped nature are established by the current variations caused by switches, commutation in motors and generators, the operation of neon signs, etc. Owing to the heavy damping, the interference is not confined to a single frequency but is distributed to some extent over the entire spectrum as in the case of atmospherics. Some of the energy is carried to the vicinity of the receiving aerial by true radiation, and where any electric circuits, e.g., the lighting or power mains, connect the place of origin with the receiving station, the latter is also subject to mains-borne interference. This is picked up by the aerial circuit of the receiver in the ordinary way. The third type of interference is caused by direct induction from the mains into the audio-frequency circuits of the receiver and is much more easily combated. The following sources of interference may be found at R.A.F. ground stations, the first-named being also the predominating cause of interference in aircraft.

(i) *Spark ignition systems.* This is of the directly radiated type and causes serious interference only on the higher frequencies. (Fig. 36a).

(ii) *Motors and generators.* Chiefly directly radiated, although a considerable portion is mains-borne. It is usually inappreciable beyond a range of 200 yards. Induction motors cause little interference as a rule. (Fig. 36b or c).

(iii) *Rectifiers, commutator type.* This apparatus is sometimes used for battery charging and for arc lamps (cinema, searchlight, etc.). Chiefly directly radiated, mains-borne component usually small. Very severe up to 200 yards range. (Fig. 36d).

(iv) *Road and rail control signals, telephone switching plant, electric bells, ovens, thermostatically controlled heating apparatus.* Chiefly mains-borne, range up to 100 yards. (Fig. 36e).

(v) *Lift plant.* Including motor and controller, also trailing cables. Both direct and mains-borne radiation, range up to 25 yards. (Fig. 36c and f.).

(vi) *Mercury arc rectifiers.* Chiefly audio-frequency induction, but possibly direct radiation also. Range may be considerable. (Fig. 36g).

(vii) *High tension transmission lines.* These cause directly radiated interference when corona losses are occurring at insulators. Range indeterminate.

(viii) *High frequency medical apparatus.* Causes very intense radiation and may have considerable range. Mains-borne interference extends to 300 yards or more.

The devices shown in fig. 36 have been successfully applied in many cases, those applicable in particular cases being denoted in the preceding paragraph. It will be observed that there is no simple remedy for items (vii) and (viii). In the latter, complete electro-magnetic screening of the room housing the medical equipment appears to be the only solution. In addition to fitting such interference suppressors at the source, devices such as that shown in fig. 36b, and 36c, may be found useful at the point where the supply mains enter the receiving room.

CHAPTER XI.—AMPLIFICATION

GENERAL PRINCIPLES

1. In preceding chapters reference has been made to the employment of the thermionic valve as an amplifier, and it now becomes desirable to consider this function in greater detail. In this connection the term valve must be understood to exclude the two-electrode valve or diode, which possesses no amplifying property. The latter is dependent upon the introduction into the space between cathode and anode of one or more grid-like structures which exercise a control upon the electron current without necessarily acting as electronic collectors, and therefore without the necessity for power expenditure in the input circuit. The triode, possessing only a single grid as a control electrode, is the prototype of all amplifying valves, and its use will be first discussed in order that the advantages obtained under certain conditions by the use of tetrodes, pentodes and other special types will be appreciated in due course.

Classification of amplifiers

2. (i) Amplification may be defined as the process by which either current, voltage or power may be increased without serious change of wave-form, although this definition does not hold in certain special cases, e.g. when it is desired to emphasize one component of a complex wave-form compared with other components. An amplifier is an assembly of valves and circuits by which amplification is achieved. Current amplification is of little practical importance, and it is usual to divide amplifiers into two main classes, (a) voltage amplifiers and (b) power amplifiers. A further classification is also made according to the portion of the frequency spectrum in which the amplifier is designed to operate, namely audio-frequency (A.F.) and radio-frequency (R.F.) amplifiers. The extent to which amplification is employed has been considerably increased in the last few years. Its use was at one time entirely confined to receivers, radio-frequency amplification being incorporated in order to increase the input voltage to the detector valve, and audio-frequency amplification of the rectified signal in order to increase the volume of sound produced. Of late years, however, radio-frequency amplification has been an important feature of many transmitters owing to the necessity for control of the radiated frequency, while audio-frequency amplification is also often required in the sub-modulator stages of R/T transmitters.

(ii) Amplifiers are also sometimes classified under the following headings :—

Class A.—An amplifier which is operated under conditions which ensure that the wave-form of the anode current variation is practically the same as that of the input grid-filament voltage. Under these conditions both the efficiency and power output are low. Class A amplifiers are used only for voltage amplification (both A.F. and R.F.) and audio-frequency power amplification in the output stage of R/T receivers, where absence of distortion is of greater importance than electrical efficiency.

Class B.—An amplifier which is operated with such negative grid bias that in the absence of any input signal voltage, the anode current is practically zero. The grid is then said to be biased to "cut-off point." Grid current is usually allowed to flow during a portion of the cycle. The anode current wave-form is approximately a series of half sine waves, alternate half-cycles being suppressed. These amplifiers are frequently employed for the amplification of modulated radio-frequency voltages in R/T transmitters ; when adjusted for maximum output the efficiency is about 40 per cent.

Class C.—In this type of amplifier the grid is biased to a point more negative than the cut-off voltage and therefore, when an input signal voltage is applied, anode current flows for less than one half-period. An efficiency of the order of 85 per cent. can be achieved in this manner. Class C amplifiers are used in both R/T and W/T transmitters. The simple C.W. transmitter

CHAPTER XI.—PARAS. 3-4

described in Chapter IX is merely a special form of R.F. amplifier in which the grid excitation voltage is derived from the output of the valve. If operated under sinusoidal conditions without grid current, the conditions are those of class A amplification, while if operated at high efficiency they correspond to class B or class C according to the magnitude of the grid bias.

Fractional band width

3. In referring to the range of frequencies over which an amplifier is designed to operate, the terms "wide" and "narrow" bands are frequently employed. The sense in which these terms are used calls for some explanation. For example, a frequency band of 20 kilocycles per second is a wide band when dealing with audio-frequency amplification, but a narrow band so far as radio frequencies are concerned. The difficulty is removed if the term "band width" is understood to signify "fractional band width". If an amplifier operates uniformly over all frequencies between an upper limit f_2 and a lower limit f_1 , it is said to respond to an absolute frequency band of width $f_2 - f_1$. The fractional band width is the quantity obtained by expressing the absolute

width as a fraction of the geometric mean frequency, $\sqrt{f_1 f_2}$, and is therefore $\frac{f_2 - f_1}{\sqrt{f_1 f_2}}$. As an

example, consider an amplifier which gives equal amplification of all frequencies between 32 and 20,000 cycles per second. The geometric mean frequency is 800 cycles per second and the fractional width 24.96, i.e. greater than unity. In contrast, take an amplifier designed to operate in the region of one megacycle per second, but which gives no appreciable amplification of frequencies 10,000 cycles above or below 1 Mc/s. The geometric mean frequency is practically

equal to 1 Mc/s and the fractional band width of the amplifier is approximately $\frac{20,000}{1,000,000} = .02$.

It will be noted that although the absolute band width is practically the same in each case (20,000 cycles per second) the fractional band width in the former example is very much greater

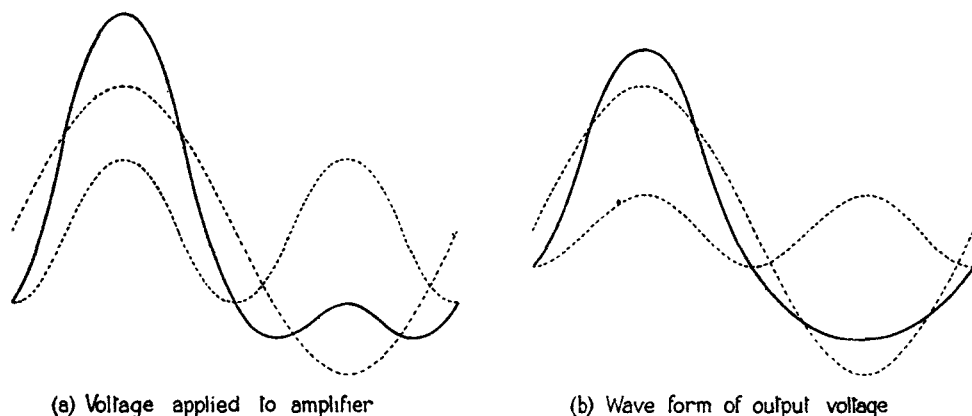


FIG. 1, CHAP. XI.—Frequency distortion.

than in the latter. When the fractional band width is small, the amplifier is said to be selective and vice versa. As a rule, radio-frequency amplifiers are designed to be as selective as possible, while audio-frequency amplifiers are required to give equal amplification over a wide (fractional) band and are therefore unselective.

Distortion in amplifiers

4. An amplifier producing an output voltage which is an exact duplicate of the input voltage in every respect save magnitude may be called an ideal or distortionless amplifier; although it

is not possible to construct such an amplifier a close approach to the ideal can be achieved by careful design. It is usual to distinguish between three types of distortion, namely :—

(i) *Frequency distortion*.—This term is applied to the unequal amplification of different frequencies. This is generally a desirable characteristic of a radio-frequency amplifier because it is then selective. In audio-frequency amplifiers, frequency distortion is usually undesirable, and its avoidance calls for considerable skill in design. The effect of this form of distortion is shown in fig. 1a, in which the original wave form has a second harmonic of amplitude one-half that of the fundamental. After amplification by a selective amplifier the wave-form may be that of fig. 1b, in which the amplitude of the second harmonic is only one-quarter of the fundamental.

(ii) *Amplitude distortion*.—The effect of amplitude distortion is illustrated in fig. 2 ; in this instance the input wave-form is assumed to be sinusoidal. After amplification the wave-form has become peaky and in fact contains second and third harmonics as well as the original wave-form. From figs. 1 and 2 it is seen that the effect of amplitude and frequency distortion may be very similar, at any rate in audio-frequency amplifiers. The difference lies rather in the cause

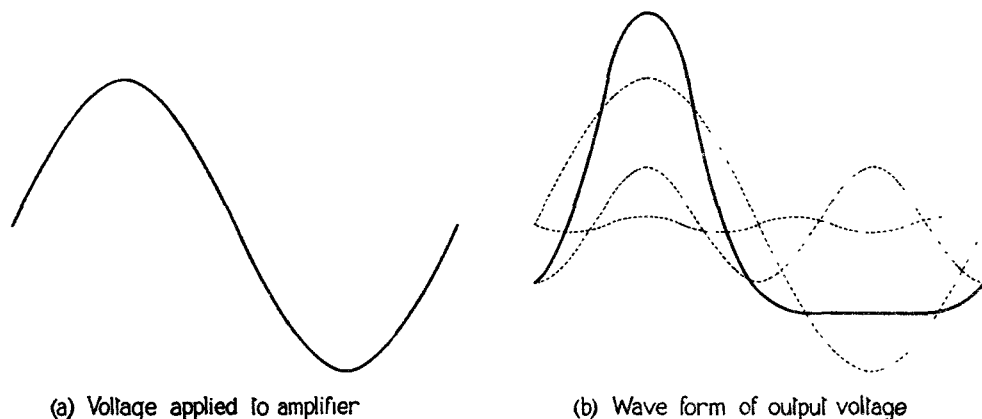


FIG. 2, CHAP. XI.—Amplitude distortion.

than in the effect, and as a rule frequency distortion is caused by the nature of the circuits used in conjunction with the valve and amplitude distortion by the valve itself. Amplitude distortion is in fact the result of a non-linear relation between current and voltage, and must exist to some extent in all valve circuits, because neither the $I_a - V_g$ nor the $I_g - V_g$ characteristic is perfectly straight. Curvature of the $I_g - V_g$ curve is immaterial if the operating conditions are so chosen that grid current never occurs. This entails operating with considerable negative bias, and restricting the amplitude of the input voltage so that the grid potential never reaches the value at which grid current commences. The use of a high effective resistance in the anode circuit tends to reduce the curvature of the dynamic $I_a - V_g$ characteristic and therefore to reduce amplitude distortion. The steps taken to minimize amplitude distortion may therefore be summarized as below.

- (a) Employ an anode circuit of high dynamic resistance.
- (b) Apply sufficient H.T. voltage to ensure that an ample approximately straight portion of the $I_a - V_g$ curve exists, in the region of negative grid voltage.
- (c) Adjust the grid bias to a value midway between the point at which the curvature of the characteristic becomes appreciable, and the point at which grid current starts to flow.
- (d) Limit the input voltage so that the excursion of anode current is confined to the straight part of the curve.

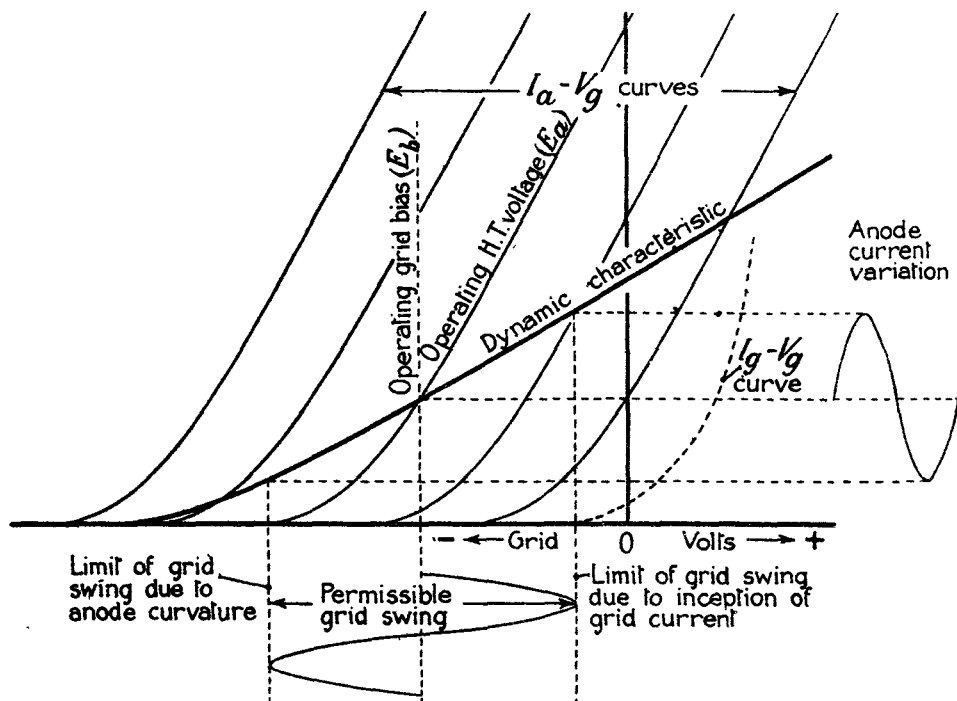


FIG. 3, CHAP. XI.—Operating conditions for distortionless amplification.

These operating conditions are shown in fig. 3. When they are fulfilled the wave-form of the anode current variation is practically identical with that of the grid-filament voltage variation. The class A amplifier may therefore be defined as one operated as indicated in (a), (b), (c) and (d) above.

(iii) *Phase distortion* results when the phase relationship between different frequency components is disturbed in such a manner that the wave-form of the output voltage differs from that of the input; although the relative amplitude of the various components is unchanged. The effect is illustrated in fig. 4, the original and distorted wave-form being again shown. The

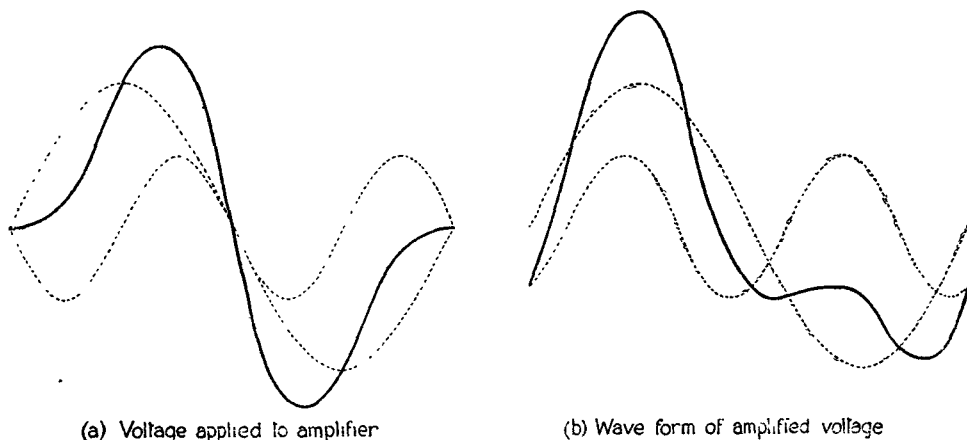


FIG. 4, CHAP. XI.—Phase distortion.

amplitudes of fundamental and second harmonic are as 2 : 1 in each case, but during amplification their relative phase has undergone a displacement of 90° . Phase distortion only occurs when the time taken for the signal to pass through the amplifier is comparable with the duration of the signal, and is therefore of no significance in ordinary reception.

Equivalent circuit of amplifier

5. The basic circuit of the triode amplifier is given in fig. 5a in which T is the triode, v_g the voltage of the signal to be amplified, E_b the voltage of the grid bias battery and E_a the voltage of the anode supply or H.T. battery. Z is the anode circuit load impedance, and for brevity is often referred to as the output circuit, while the circuit connected between grid and filament of the valve is called the input circuit. The variation of grid-filament voltage v_g gives rise to a corresponding variation of anode current, which must of necessity flow through the load impedance, and consequently a varying voltage v_a is set up across the latter. In a voltage amplifier the object is to obtain the largest possible voltage variation v_a . In a power amplifier, however, in addition to this voltage variation, an appreciable current variation is also required, so that the power dissipated in the output circuit, as a result of the grid-filament voltage variation v_g , shall be as large as possible. The variation of anode current produced by the application of a voltage v_g to the grid-filament path is exactly the same as would be produced by a voltage μv_g acting in

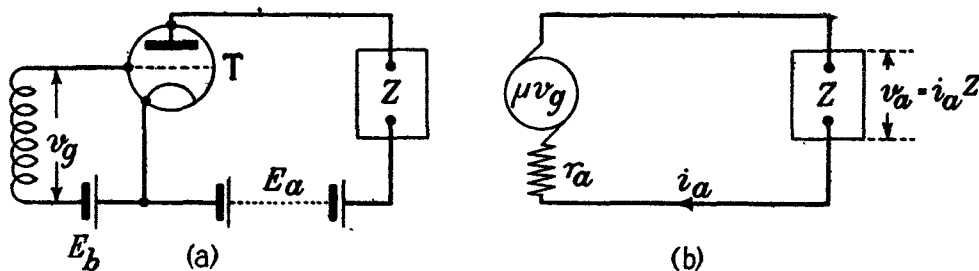


FIG. 5, CHAP. XI.—Basic circuit of amplifier, and equivalent circuit.

the anode circuit of the valve (Chap. VIII). The equivalent circuit of the valve amplifier is therefore as shown in fig. 5b. This equivalent circuit gives only those currents and P.D.'s which result from the application of the signal voltage, which are superimposed upon the steady or no-signal values of P.D. and current.

Voltage amplification factor

6. (i) The voltage amplification factor (V.A.F.) of a valve and an associated impedance in its anode circuit is the ratio of the voltage variation across the external impedance to the voltage variation between grid and filament of the valve. The anode circuit load impedance may be of any nature whatever, provided that it possesses finite conductivity at zero frequency, i.e. for direct current. This limitation merely signifies that a simple series condenser cannot be employed, because it is necessary to provide a complete conductive path for the steady component of anode current. In practice the anode load impedance may be a resistance, an inductive choke, or a tuned circuit consisting of inductance and capacitance in parallel. Further, an additional circuit may be coupled to the load impedance by any of the methods enumerated in Chapter VI.

(ii) In deriving the voltage amplification factor appropriate to any particular form of anode circuit, it is desirable to assume that no amplitude distortion will take place, and therefore that the following conditions are fulfilled.

(a) Ample filament emission is provided.

(b) The anode is maintained at a positive potential with respect to the filament during the whole cycle of applied grid voltage.

CHAPTER XI.—PARA. 7

(c) The mean grid potential is sufficiently negative to ensure that no grid current flows.

(d) The $I_a - V_g$ characteristic is approximately straight over the whole portion to which the excursion of anode current extends.

(e) Unless the contrary is explicitly stated, the interelectrode capacitance is assumed to be negligibly small; the effect of its finite magnitude will be indicated where this is of consequence in operation or design.

Ohmic resistance as load impedance

7. In the circuit shown in fig. 6a the load impedance consists of a non-inductive resistance of R ohms. If an alternating voltage $v_g = \mathcal{V}_g \sin \omega t$ is applied between grid and filament, its effect upon the anode current is exactly the same as would be produced if the grid potential

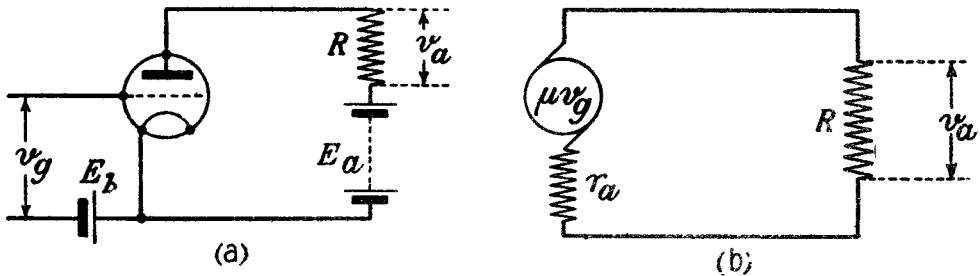


FIG. 6, CHAP. XI.—Resistance as anode load.

were maintained constant and an alternating E.M.F. $\mu \mathcal{V}_g \sin \omega t$ were introduced into the anode circuit (fig. 6b). The resulting variation of anode current will be

$$i_a = \frac{\mu \mathcal{V}_g \sin \omega t}{r_a + R}$$

The output voltage, v_a , of the amplifier is the P.D. across the load resistance R and is equal to $i_a R$, hence

$$v_a = \frac{R \mu \mathcal{V}_g \sin \omega t}{r_a + R}$$

With a sinusoidal variation of grid voltage, therefore, the output voltage v_a is also sinusoidal and the above equation may be written

$$v_a = \frac{R}{r_a + R} \mu v_g.$$

The V.A.F. is defined above as the ratio $\frac{v_a}{v_g}$ so that the V.A.F. of the circuit shown in fig. 6 is

$$\frac{v_a}{v_g} = \frac{R}{r_a + R} \mu.$$

Example.—A triode has an amplification factor of 20 and an r_a of 50,000 ohms. If a purely resistive impedance of 100,000 ohms is placed in the anode circuit, find the V.A.F.

$$\begin{aligned} \text{V.A.F.} &= \frac{R}{r_a + R} \mu \\ &= \frac{100,000}{50,000 + 100,000} \times 20 \\ &= 13\frac{1}{3}. \end{aligned}$$

8. An increase in the value of the load resistance will result in an increase in the V.A.F. ; for example if a resistance of 200,000 ohms is substituted in the circuit just considered, a V.A.F. of 16 is obtained. The limiting value of the V.A.F. is equal to the amplification factor of the valve, i.e. μ , but this value cannot be reached in practice. It is approached more closely as the ratio $\frac{R}{r_a}$ becomes larger and larger, as shown in fig. 7, in which the ordinate is $\frac{\text{V.A.F.}}{\mu}$ and the abscissa the ratio $\frac{R}{r_a}$, so that to obtain the V.A.F. with any particular valve, the ordinate must be multiplied by the appropriate value of μ . It may appear desirable to use the highest obtainable value of load resistance ; a practical limit to its magnitude is however imposed by the necessity of maintaining the anode at a positive potential of at least a few volts (e.g. 10 volts) above that of the filament. Since the anode D.C. resistance of the valve and the load resistance are in series, this limits the value of the latter to about $10 r_a$. Reference to fig. 7 shows that the V.A.F. is then

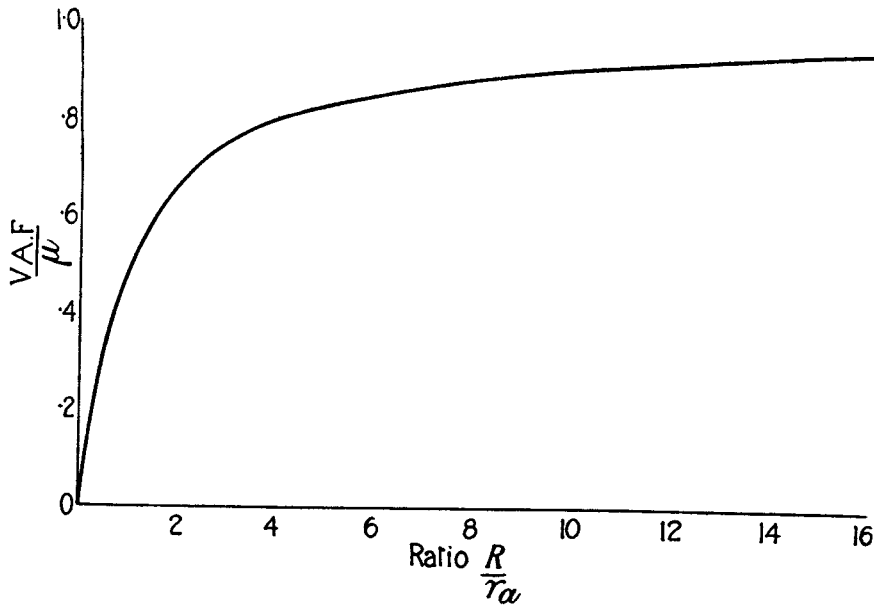


FIG. 7, CHAP. XI.—Effect of ratio $\frac{R}{r_a}$ upon V.A.F.

90 per cent. of the theoretical maximum value μ . From the expression given above it would also appear that the V.A.F. is independent of the frequency of the applied grid-filament voltage. In obtaining the V.A.F., however, the presence of stray capacitance was neglected, and it will presently be shown that the effect of such a capacitance, which effectively acts as a shunt upon the load resistance, is to cause serious reduction of the amplification obtainable at all frequencies above a few thousand cycles per second.

Inductance as load impedance

9. Fig. 8a shows a possible amplifier disposition in which the impedance Z of fig. 5 is constituted by an inductance of L henries and of negligible resistance, the equivalent circuit being given in fig. 8b. The mean operating potentials E_b and E_a are as in the previous example.

CHAPTER XI. — PARA. 9

The application of a sinusoidal grid-filament voltage $v_g = \mathcal{V}_g \sin \omega t$ will result in a corresponding variation of anode current i_a , and

$$i_a = \frac{\mu v_g}{\sqrt{r_a^2 + (\omega L)^2}}$$

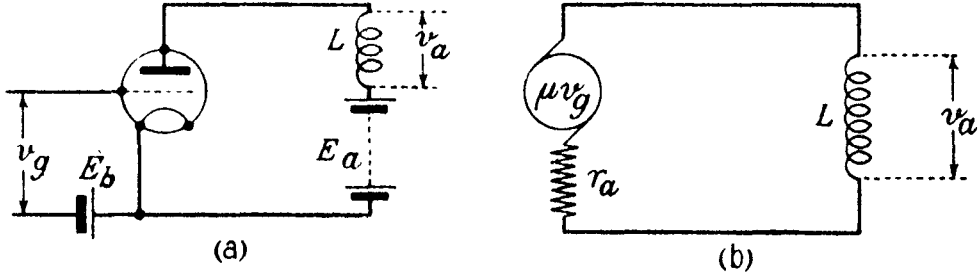


FIG. 8, CHAP. XI. Inductance as anode load.

The P.D. between the ends of the inductance, which is of course the output voltage v_a of the amplifier is $\omega L i_a$, or

$$v_a = \frac{\omega L}{\sqrt{r_a^2 + (\omega L)^2}} \mu v_g$$

and the V.A.F. is

$$\frac{v_a}{v_g} = \frac{\omega L}{\sqrt{r_a^2 + (\omega L)^2}} \mu$$

Example.—In fig. 8a, if the valve has an amplification factor of 20 and an r_a of 20,000, while the load impedance is an inductance of 10 henries, find the V.A.F. when the frequency of the applied grid filament voltage is (i) 10 cycles per second, (ii) 100 cycles per second, (iii) 1,000 cycles per second.

$$(a) \omega L = 2\pi \times 10 \times 10 = 200\pi = 628.$$

$$\text{V.A.F.} = \frac{628}{\sqrt{20,000^2 + 628^2}} \times 20$$

$$\doteq \frac{628}{20,000} \times 20$$

$$\doteq 0.628$$

$$(b) \omega L = 6,280.$$

$$\text{V.A.F.} = \frac{6,280}{\sqrt{20,000^2 + 6,280^2}} \times 20$$

$$\doteq \frac{6,280}{21,000} \times 20$$

$$\doteq 6.6$$

$$(c) \omega L = 62,800.$$

$$\text{V.A.F.} = \frac{62,800}{\sqrt{20,000^2 + 62,800^2}} \times 20$$

$$\doteq \frac{62,800}{66,400} \times 20$$

$$\doteq 18.95.$$

With this form of load impedance, then, the V.A.F. increases as the frequency of the input voltage is increased, and approaches the limiting value μ as the load reactance, ωL , becomes larger and larger compared with the anode A.C. resistance r_a . The increase of V.A.F. obtained by an increase in the ratio $\frac{\omega L}{r_a}$ is shown in fig. 9 which should be compared with fig. 7. It is seen that no advantage is obtained by increasing the ratio $\frac{\omega L}{r_a}$ beyond about 5, ω being taken as 2π times the lowest frequency at which appreciable amplification is required. It should be noted that as the inductance is assumed to have zero resistance, there is no steady voltage drop between

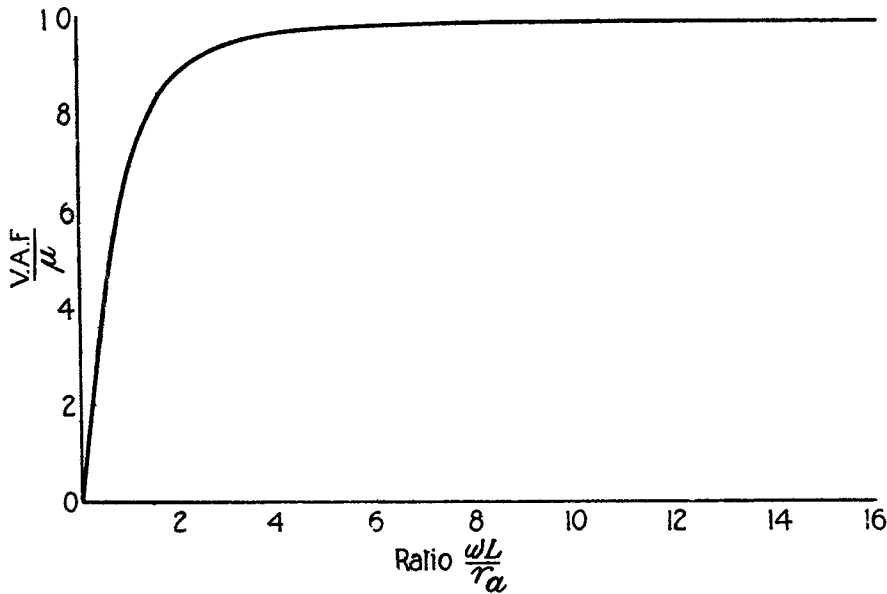


FIG. 9, CHAP. XI. Effect of ratio $\frac{\omega L}{r_a}$ upon V.A.F.

its'ends, and the mean anode-filament P.D. is equal to the E.M.F. of the H.T. battery. In practice the resistance of the coil is always negligible compared with the anode-filament (D.C.) resistance of the valve.

Effect of stray capacitance

10. Before discussing the practical application of these and other forms of load impedance, it is desirable to study the effect of a stray capacitance in parallel with the inductance or resistance constituting the load, such a capacitance is always present, and its effective value may be of the order of $100 \mu\mu F$. Let the stray capacitance be denoted by C_s , its reactance at a

frequency $\frac{\omega}{2\pi}$ being $\frac{1}{\omega C_s}$. Taking the case of a resistance load, the anode circuit impedance,

CHAPTER XI.—PARA. 10

Z_o is then that of R and C_s in parallel. By methods explained in Chap. V it can be shown that

$$Z_o = \frac{R}{\sqrt{1 + \omega^2 C_s^2 R^2}} = \sqrt{X_o^2 + R_o^2}$$

$$X_o = \frac{\omega C_s R^2}{1 + (\omega C_s R)^2}$$

$$R_o = \frac{R}{1 + (\omega C_s R)^2}$$

and the V.A.F. becomes

$$\frac{Z_o}{\sqrt{(r_a + R_o)^2 + X_o^2}} \mu.$$

Example.—Assuming that $r_a = 50,000$ ohms, $\mu = 20$, $R = 100,000$ ohms, $C_s = 100 \mu F$, calculate the V.A.F. at the following frequencies, viz., (i) 800 cycles per second (ii) 16,000 cycles per second (iii) 800 kilocycles per second, the first being near the mean audio frequency, the second near the upper limit of audibility, and the third a radio frequency in the middle of the medium broadcast band.

At 800 cycles per second,

$$\omega = 5,000 \text{ (approx.)}$$

$$\omega C_s R = 5 \times 10^3 \times 10^2 \times 10^{-12} \times 10^5 = .05$$

$$(\omega C_s R)^2 = .0025$$

$$X_o = \frac{.05 \times 10^5}{1.0025} = 4,988 \text{ ohms}$$

$$R_o = \frac{10^5}{1.0025} = 99,750 \text{ ohms}$$

$$Z_o = 99,875 \text{ ohms}$$

$$\text{V.A.F.} = \frac{99,875}{\sqrt{(5 \times 10^4 + 9.975 \times 10^4)^2 + 4,988^2}} \times 20$$

As X_o is so small compared to $r_a + R_o$ this is practically

$$\frac{9.9875}{14.975} \times 20$$

$$\text{V.A.F.} = 13.3.$$

This result is of course identical with that found by assuming the stray capacitance to be negligible.

At 16,000 cycles per second,

$$\omega = 100,000$$

$$\omega C_s R = 10^5 \times 10^2 \times 10^{-12} \times 10^5 = 1$$

$$(\omega C_s R)^2 = 1$$

$$X_o = \frac{10^5}{2} = 50,000 \text{ ohms}$$

$$R_o = \frac{10^5}{2} = 50,000 \text{ ohms}$$

$$Z_o = \sqrt{50,000^2 + 50,000^2}$$

$$= 70,700 \text{ ohms}$$

$$\text{V.A.F.} = \frac{70,700}{\sqrt{(50,000 + 50,000)^2 + 50,000^2}} \times 20$$

$$= \frac{7.07}{125} \times 20$$

$$= 12.6.$$

At 800 kilocycles/sec.

$$\begin{aligned}
 \omega &= 5 \times 10^5 \\
 \omega C_s R &= 5 \times 10^5 \times 100 \times 10^{-12} \times 10^5 = 50 \\
 (\omega C_s R)^2 &= 2,500 \\
 X_o &= \frac{50 \times 10^5}{2,501} \div \frac{50 \times 10^5}{2,500} \div 2,000 \\
 R_o &\div \frac{10^5}{2,501} \div \frac{10^5}{2,500} \div 40 \\
 \therefore Z_o &= \sqrt{R_o^2 + X_o^2} \div 2,000 \\
 \text{V.A.F.} &= \frac{2,000}{\sqrt{50,040^2 + 2,000^2}} \times 20 \\
 &\div \frac{2,000}{50,000} \times 20 \\
 &\div .8.
 \end{aligned}$$

11. From the above example, the following general results may be deduced. The effect of stray capacitance upon the V.A.F. is negligible at the lower audio frequencies, but becomes of some importance at the higher audio frequencies. At that frequency for which the reactance of the stray capacitance $\left(\frac{1}{\omega C_s}\right)$ is equal to the joint resistance of the valve and load resistance in parallel, i.e. when

$$\frac{1}{\omega C_s} = \frac{R r_a}{R + r_a},$$

the V.A.F. is reduced to $.707\mu$, and falls off rapidly when the frequency is further increased. With the valve and circuit specified above, the corresponding frequency is about 480 kc/s. It may be taken as a general rule that with the valves at present employed, a resistive anode load will give no appreciable amplification at frequencies higher than about 500 kc/s.

Development of tuned anode circuit

12. (i) Now consider the effect of a similar capacitance in the case of the amplifier having an inductive anode load. It is at once apparent from the circuit diagram of fig. 10 that the capacitance and inductance together form a tuned circuit, the resonant frequency of which is

equal to $\frac{1}{2\pi\sqrt{LC_s}}$ cycles per second. If the resistance were truly zero, the circuit would

behave at this frequency as a perfect rejector, offering an infinitely high opposition to the flow of current, and the V.A.F. would reach the theoretical limiting value μ . As however some slight resistance must exist, the opposition of the parallel circuit at its resonant frequency is

not infinite, but is equal to that of a purely resistive impedance of $\frac{L}{C_s R}$ ohms. This apparent

resistance is termed the dynamic resistance of the circuit and is denoted by R_d . (Chapt. V.) In certain circumstances the stray capacitance may be deliberately augmented by connecting a condenser of capacitance C in parallel with the inductance; the resonant frequency is then

equal to $\frac{1}{2\pi\sqrt{L(C+C_s)}}$ and the load resistance is $\frac{L}{(C+C_s)R}$ ohms. When a parallel $L C$

CHAPTER XI.—PARAS. 13-14.

circuit is employed in this manner it is called a "tuned anode" circuit. The V.A.F. follows at once from the results obtained with a purely resistive anode load, i.e.

$$\text{V.A.F.} = \frac{R_d}{r_a + R_d} \mu$$

This expression is of course only valid when the parallel circuit $L, (C + C_s)$, is tuned to the input frequency. At frequencies higher than the resonant frequency, the circuit will offer both resistance and capacitive reactance, while at frequencies below the resonant frequency it will behave as an inductive resistance, its impedance being always less than the dynamic resistance under resonant conditions. Hence an amplifier having an inductance in its anode circuit will always give greatest amplification at some particular frequency, depending upon the inductance

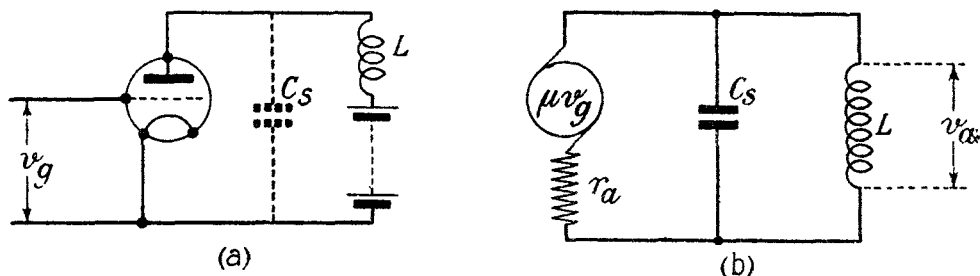


FIG. 10, CHAP. XI. Effect of stray anode-filament capacitance.

and the capacitance of the circuit, i.e. frequency distortion must exist in some degree. The amplifier will give a fairly even response over a wide frequency band, only if a large ratio of inductance to total capacitance is maintained.

(ii) The forms of load impedance just described form the basis of all amplifier designs and their application to practical receiving amplifiers will now be described.

RECEIVING AMPLIFIERS

13. The ultimate purpose of a radio receiver is to produce an audio-frequency variation of current in the windings of the telephone receiver, and a consequent emission of sound by the latter. In Chapter X it was shown that this invariably involves some form of rectification of the radio-frequency voltage, so that an audio-frequency variation of current will result. The purpose of amplification is to obtain either a greater signalling range for the same transmitter power, or a greater response, i.e. a louder sound, from the telephone receivers (for the same power and signalling range) or both. In general, amplification of the radio-frequency voltage, before rectification, will increase the signalling range, while amplification of the rectified output from the detector valve will increase the sound output from the telephones.

Radio-frequency and audio-frequency amplification

14. (i) In the early development of radio-frequency amplification it was thought desirable to aim at the production of aperiodic or semi-aperiodic amplifiers, which would amplify all frequencies, or a wide band of frequencies, equally well. The advantage gained by aperiodic amplification is the elimination of tuning controls, manipulation being confined to the adjustment of the aerial circuit and of the reaction control. The employment of a tuned amplifier of correct design endows the receiver with such enhanced selectivity that aperiodic R.F. amplification has fallen into complete disuse, and tuned radio-frequency amplifiers are now employed for the purpose of gaining selectivity, even if, in the absence of interference, the desired signalling range could be attained by a simple receiver of the type described in Chapter X.

(ii) A further advantage of radio-frequency amplification lies in the fact that for small input voltages, the output from any practical form of rectifier is approximately proportional to the square of the input voltage. If a certain small signal voltage is available, and is amplified, say, to four times its initial magnitude before being applied to the rectifier, the rectified output is sixteen times that which would be obtained by applying the signal voltage directly to the rectifier. The total R.F. amplification is however limited by certain factors to be discussed later, and in many cases the rectified signal may still be of insufficient amplitude to give the desired output of sound from the telephones. In a perfectly quiet room, a very faint telephone sound, such as a morse signal from a very distant transmitter, may prove to be quite readable, but it is necessary to increase the input to the telephone to an enormous degree in order to read a signal in noisy surroundings. In the reception of radio signals in an aeroplane, the high noise level renders it desirable to use the strongest signal which the telephone receivers will withstand without overload, and recourse to a considerable degree of audio-frequency amplification is necessary. As a rule, the audio-frequency amplifier is required to deal with a considerable frequency range, and should amplify equally well at all frequencies within this range, although for C.W. reception, a sharply tuned audio-frequency amplifier, or note selector, is sometimes employed. The actual volume of sound depends upon the supply of electric power to the telephone receivers and the final stage in the receiver must operate as a power amplifier. It must be noted that the term "power amplifier" carries no implication as to the amount of power supplied to the sound-producing device but is merely an indication of its mode of operation.

Multi-stage amplifiers

15. (i) Summarizing the above, then, it may be said that a typical radio receiver may consist of four portions, executing the following functions :—

(a) Radio-frequency voltage amplification, in order to ensure the maximum input grid voltage to the detector, and to achieve the highest possible degree of selectivity.

(b) Rectification or detection, which gives an audio-frequency output in response to a radio-frequency input voltage.

(c) Audio-frequency voltage amplification, by which the audio-frequency output voltage of the detector is increased in amplitude.

(d) Power amplification, giving the maximum transfer of power from the H.T. supply device to the telephone receiver, for a given input voltage from the preceding stage.

In addition, some form of heterodyne will be necessary for C.W. reception. This may be incorporated in the receiver either by the employment of sufficient reaction at some point in the radio-frequency portion (autodyne) or by an in-built separate heterodyne. Alternatively, an external heterodyne such as the syntonizer may be employed as described in the previous chapter.

(ii) In such a receiver each valve and its anode circuit is generally called a "stage" : for example, a receiver may have two radio-frequency amplifying stages, a detector stage, a stage of audio-frequency voltage amplification and a final power amplifying stage. It is usual to arrange that all the valve filaments of a multi-stage receiver are supplied from a single source, a common grid bias supply and common H.T. supply being also adopted. If these sources are either primary or secondary batteries the receiver is said to be "battery operated," while if arrangements are made to enable the supplies to be drawn from the power mains (either D.C. or A.C.) the receiver is said to be "mains operated." In the following discussion, battery operation is assumed unless otherwise stated ; mains operation is usually employed when a considerable sound output is required, as in public address systems and in broadcast receivers. In multi-stage amplifiers, the output circuit of one valve is the input circuit of its successor, and the valves are said to be coupled together owing to this dual function. Care must be taken not to confuse the term "inter-valve coupling" with the idea of coupling between tuned circuits as a means of transferring energy from one to the other. The purpose of inter-valve coupling is merely to apply the voltage developed by one valve to the grid and filament of the next, and energy transfer is generally to be avoided.

CHAPTER XI.—PARAS. 16-17

Audio-frequency amplification

16. Three forms of inter-valve coupling are in general use in A.F. circuits. These are :—

- (i) Resistance-capacitance coupling.
- (ii) Choke-capacitance coupling.
- (iii) Transformer coupling.
 - (a) Series feed.
 - (b) Parallel feed.

Resistance capacitance coupling

17. (i) Referring to fig. 11a, when an alternating voltage v_{g1} is applied to the grid and filament of the valve T_1 , an alternating voltage v_a is developed across the ends of the resistance R . It is desired to apply this voltage to the grid and filament of the following valve T_2 . Now the upper end of the resistance is already connected to the filament of the valve T_2 through the H.T. battery, so that it is apparently only necessary to connect the anode of the valve T_1 to the grid of T_2 . This connection has been inserted in the diagram as a dotted line. Unfortunately, the addition of such a connection would place the grid of T_2 at a positive potential, equal to the E.M.F. of the H.T. battery, with respect to its filament, a heavy grid current would therefore flow, and the valve would probably be destroyed; it would at least fail to function in the desired manner.

(ii) This effect is avoided by the insertion of a condenser C_g into the anode-grid connecting lead (fig. 11b). Under these conditions the grid is completely insulated from the filament, a

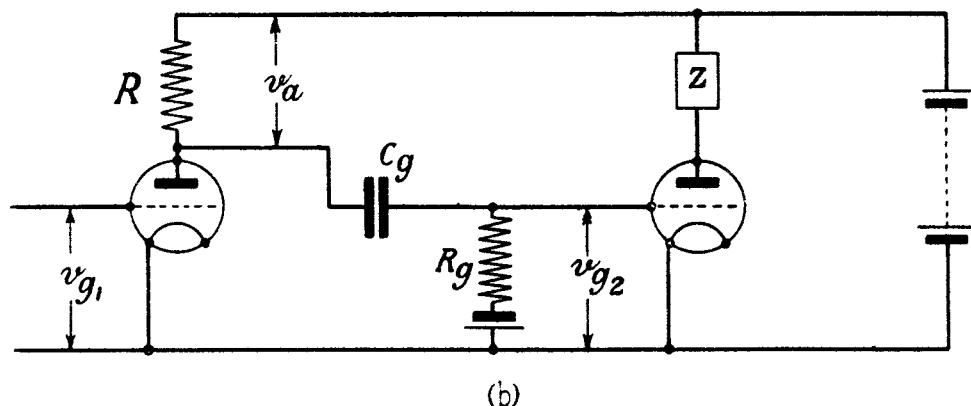
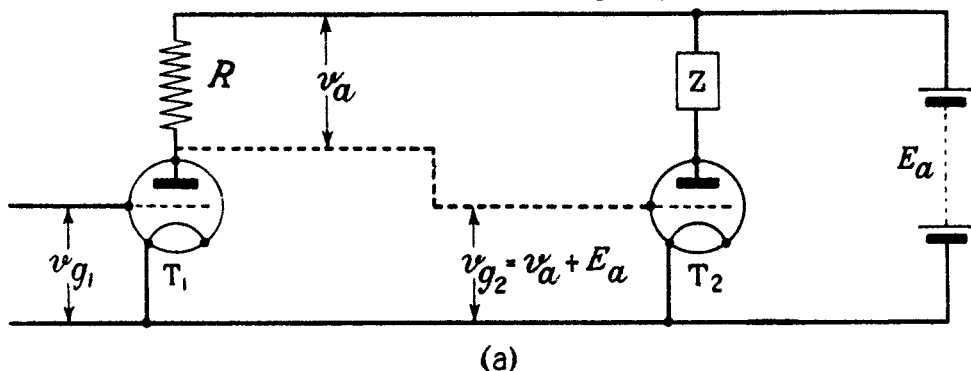


FIG. 11, CHAP. XI. Development of resistance-capacitance coupling.

condition which is not permissible, for the grid will inevitably collect a few electrons during each successive positive half-cycle of input voltage. Unless it is possible for these to return to the filament, the grid will assume an increasingly negative potential, ultimately reducing the anode current to zero. This phenomenon has already been referred to with regard to both detector and oscillator circuits, and the remedy in the present case is the same, namely the addition of a grid leak resistance R_g , which must be externally connected between the grid and filament of the valve. Any required biasing voltage may be applied by connecting a battery of the required voltage and polarity in series with the grid leak. In order that the voltage applied to the grid and filament of the succeeding valve shall be as large as possible, the grid leak should have a high resistance, e.g. $\cdot 1$ to 1 megohm, and the capacitance of the grid condenser should also be large so that its reactance at the operating frequency is small compared with the resistance of the leak. As will be seen later, the values of C_g and R_g are chosen with regard to these and certain other considerations.

Stage gain

18. The "gain" of a stage of amplification is the ratio of the grid-filament signal P.D. v_{g2} of one valve to the input (grid-filament) voltage v_{g1} of the previous valve in the amplifier, and differs from the V.A.F. in that any voltage drop in the inter-valve coupling device is taken into consideration. In the present type of circuit the impedance of the grid condenser and leak (in series), forms a shunt upon the anode resistance R , and the V.A.F. is rather less than that calculated by the formula given above. At frequencies for which the reactance of the grid condenser is small compared with the resistance of the grid leak, the V.A.F. may be calculated approximately by assuming that the load resistance R_e consists of R and R_g in parallel, i.e.

$$R_e = \frac{R_g}{R + R_g} R.$$

The factor $\frac{R_g}{R + R_g}$ is obviously only of importance when R_g is not very much larger than R .

Its effect is only mentioned because it forms one limit to the upper value which may usefully be adopted for the resistance R ; as R_g will generally be not more than a megohm, the effective load resistance will be less than this even if R is very much greater. The stage gain is rather less than the V.A.F. because the grid condenser and leak together form a kind of alternating current potentiometer, supplying only a fraction of the P.D. across the anode load to the grid and filament of the succeeding valve, i.e.

$$v_{g2} = \frac{R_g}{\sqrt{R_g^2 + \left(\frac{1}{\omega C_g}\right)^2}} v_a.$$

At the frequency at which $R_g = \frac{1}{\omega C_g}$ the gain will be nearly $\frac{1}{\sqrt{2}}$ or 70 per cent. of the theoretical

V.A.F. At the frequency which makes $\frac{1}{\omega C_g} = 2R_g$ the stage gain is nearly 50 per cent. of the

V.A.F., while when the frequency is so high that $\frac{1}{\omega C_g}$ is less than $\frac{R_g}{3}$ the stage gain is within

5 per cent. of the V.A.F. The capacitance of the grid condenser depends upon the lowest frequency at which appreciable amplification is required, but should not be larger than necessary. The valves preceding the audio-frequency stages, i.e. R.F. amplifying valves and detector valve, are not absolutely steady in action but tend to set up a background of noise, which consists for the most part of very low frequency components. If the amplifier has negligible gain for frequencies below about 200 cycles per second, this background noise is reduced to a considerable extent. As an example, if we employ a grid condenser and leak of $\cdot 001 \mu F$ and 2 megohms respectively, the gain will be 70 per cent. of the V.A.F. at about 80 cycles per second,

CHAPTER XI.—PARAS. 19-20

but will fall off rapidly at lower frequencies, being only 45 per cent. of the V.A.F. at 40 cycles per second. The actual amplification of frequencies below 200 cycles per second is shown in fig. 12 as a fraction of the V.A.F. and the reduction of low frequency noise due to preceding stages is seen to be considerable.

19. The extent to which the stage gain falls off at high frequencies is determined by the magnitude of the effective shunting capacitance, compared with the anode A.C. resistance of the valve and the circuit resistances R and R_g . The lower the values of these resistances the less will be the effect of the capacitance and if it is desired to avoid considerable fall in amplification

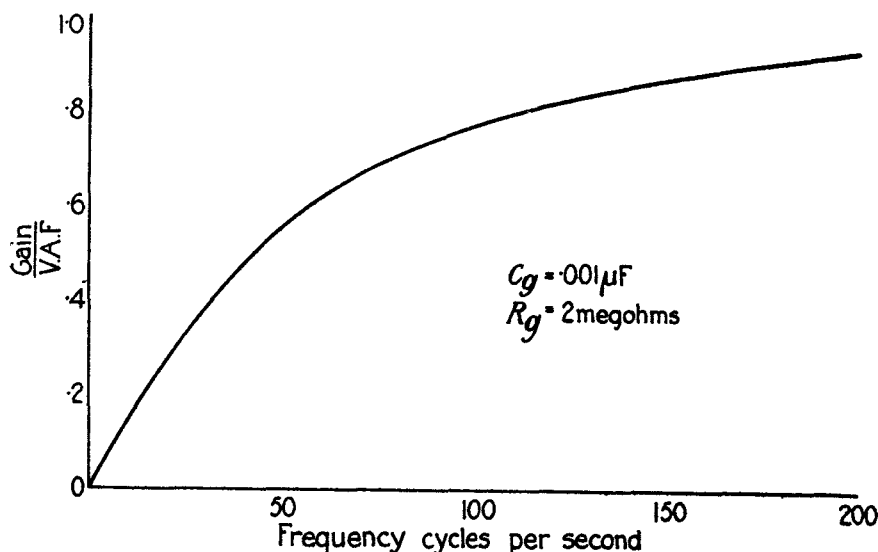


FIG. 12, CHAP. XI. Low-frequency cut-off due to grid condenser.

at the higher audio frequencies, it is necessary to employ a valve of low amplification factor, because such a valve usually has a low anode A.C. resistance. This in turn enables a V.A.F. approaching the limiting value μ to be achieved with a comparatively low anode resistance R and a smaller grid leak resistance than would otherwise be required. The effective shunt capacitance is approximately equal to $C_s + C_{af} + C_{gt} + C_{ag}(1 + A)$ where

C_s = stray capacitance between anode and filament of first valve.

C_{af} = anode-filament capacitance of first valve.

C_{gt} = grid-filament capacitance of second valve.

C_{ag} = anode-grid capacitance of second valve.

A = V.A.F. of the succeeding valve and its associated anode circuit.

Note.— A cannot exceed the amplification factor μ of the second valve.

Since each interelectrode capacitance (including that due to the valve holder) usually ranges from 5 to 10 $\mu\mu F$, it will be seen that the previous estimate of 100 $\mu\mu F$ for the total effective shunt capacitance is a fair average value. The effect of the last term, i.e. $C_{ag}(1 + A)$ is important, because it renders the gain of one stage dependent upon that of the succeeding valve. The overall gain of a number of amplifying stages can, therefore, only be accurately calculated by commencing with the final stage and working backwards, allowing for the effective load due to the shunting capacitance in each case.

20. An important practical point in the design and maintenance of this type of amplifier is the insulation resistance of the coupling condenser. Suppose that in fig. 11b

the insulation resistance of C_g is 20 megohms. If the anode resistance R is .2 megohm and the grid leak 2 megohms, the H.T. battery voltage being say 222 volts, a direct current I_g will flow in the path R, C_g, R_g , its value being $\frac{222}{22 \cdot 2 \times 10^6}$ amperes or 10 microamperes. The P.D. between the ends of the grid leak will be $R_g I_g = (.2 \times 10^6) (10 \times 10^{-6}) = 2$ volts, and the grid will be positive with respect to the filament by this amount. Hence the grid bias voltage actually applied must be two volts greater than that indicated by a consideration of the $I_a - V_g$ curve of the valve. If the insulation resistance falls below the value given, an even greater positive bias will be applied to the grid, and this must be neutralized by the application of opposite bias. In practice, therefore, the insulation resistance of the grid condenser must be maintained at a high value, and moisture, dirt or dust will inevitably lead to a reduction in the amplification. Only high quality mica dielectric is suitable for grid condensers.

Choke-capacitance coupling

21. (i) Instead of placing a resistance in the anode circuit in order to develop the amplified voltage, an inductive coil may be used, as already shown. The coupling to the succeeding valve is made by means of a grid condenser as in the case of resistance-capacitance coupling. It has already been demonstrated that at the lower audio frequencies the effect of the shunt capacitance is negligible and the V.A.F. is practically equal to

$$\frac{v_a}{v_{g1}} = \frac{\omega L}{\sqrt{r_a^2 + (\omega L)^2}} \mu$$

The insertion of the grid condenser and leak will cause the stage gain to be less than this, for those frequencies at which the reactance of the grid condenser is comparable with the resistance of the leak, so that the gain is

$$\frac{v_{g2}}{v_{g1}} = \frac{R_g}{\sqrt{R_g^2 + \left(\frac{1}{\omega C_g}\right)^2}} \times \frac{\omega L \mu}{\sqrt{r_a^2 + (\omega L)^2}}, \text{ approximately.}$$

In practice, the inductance of the choke is so chosen that in conjunction with its own self-capacitance and the effective shunt capacitance previously alluded to, the anode circuit is a rejector for some frequency in the middle portion of the audio-frequency range, say 1,000 cycles per second. Allowing say 100 $\mu\mu F$ for the self-capacitance of the coil, the total capacitance may be of the order of 200 $\mu\mu F$. The inductance required to tune to 1,000 cycles per second may now be found from the formula

$$f = \frac{1}{2\pi \sqrt{L C}},$$

$$\text{i.e. } L = \frac{1}{2^2 \pi^2 f^2 C}. \quad (\text{N.B. } 2^2 \pi^2 \doteq 40)$$

$$\text{or } L = \frac{10^{12}}{40 \times 10^6 \times 200} = 125 \text{ henries.}$$

At higher frequencies the amplification will fall off owing to the effect of the shunt capacitance and the overall response of the amplifier will be somewhat as shown in fig. 13.

(ii) This form of coupling is superior to resistance-capacitance coupling in that a somewhat higher stage gain can be achieved, and also because the steady voltage drop in the inductance is negligible, so that a lower H.T. supply voltage may be used. These advantages are however completely offset by the weight and cost of the inductance, and the serious reduction of amplification which occurs at the higher and lower ends of the audio-frequency range. Resistance-capacitance coupling is therefore generally preferred for service use, except where any particular stage is required to act as a note selector.

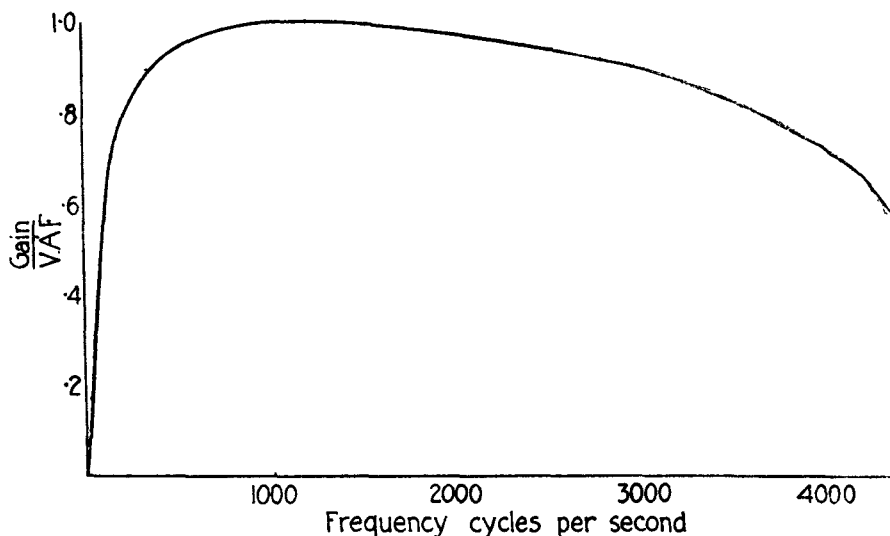


FIG. 13, CHAP. XI. Frequency response of choke-capacitance coupled amplifier.

Note selector

22. In C.W. reception the heterodyne beat note is under the control of the operator, and a high degree of selectivity can be attained by the use of a sharply tuned audio-frequency stage. The selectivity of a choke-capacitance coupled stage depends chiefly upon the design of the anode circuit impedance, an even response being obtained by a high $\frac{L}{C}$ ratio ; in a selective amplifier the inductance is reduced and the capacitance increased, the product LC being so chosen that the anode circuit is resonant to the frequency at which the telephone receivers give maximum response, i.e. about 1,000 cycles per second. The circuit diagram is given in fig. 14. This is so arranged that when high audio-frequency selectivity is not desired, e.g. for R/T reception, the tuned circuit can be switched out and a resistance R_0 substituted.

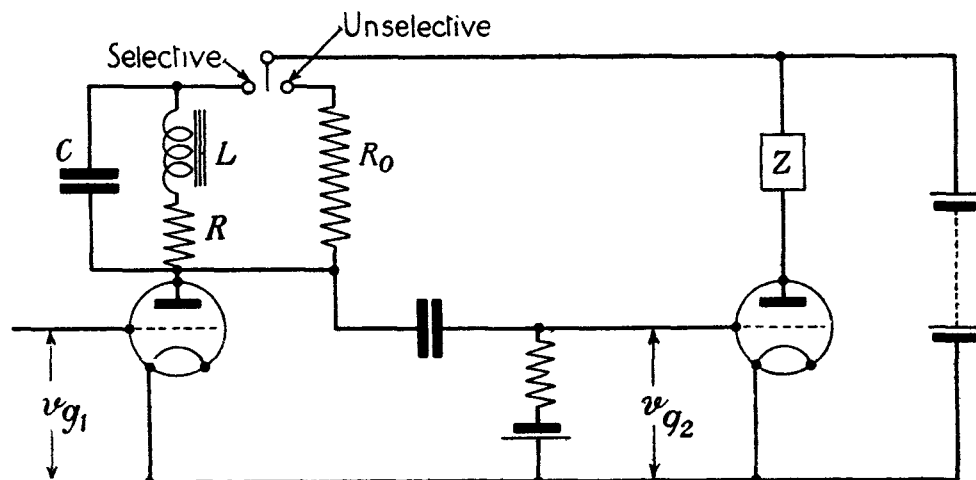


FIG. 14, CHAP. XI.—Note selector.

23. An expression showing the V.A.F. of a "tuned anode" amplifier at any frequency $\frac{\omega}{2\pi}$ may be derived as follows. The impedance operator of the inductive branch of the anode circuit is $R + j\omega L$, while that of the capacitive branch is $\frac{1}{j\omega C}$. The impedance operator of these branches in parallel is

$$z = \frac{\frac{1}{j\omega C} (R + j\omega L)}{R + j \left(\omega L - \frac{1}{\omega C} \right)}$$

If R is small compared with ωL this approximates closely to

$$z = \frac{L}{C \left\{ R + j \left(\omega L - \frac{1}{\omega C} \right) \right\}}$$

The current established in the anode circuit by a sinusoidal voltage v_{g1} is, in the notation of paragraph 64, Chapter V,

$$i_a = \frac{\frac{\mu}{L}}{r_a + \frac{L}{C \left\{ R + j \left(\omega L - \frac{1}{\omega C} \right) \right\}}} v_{g1} \quad (1)$$

and the P.D. across the output circuit is

$$\begin{aligned} v_a &= \frac{\frac{L}{C \left\{ R + j \left(\omega L - \frac{1}{\omega C} \right) \right\}}}{r_a + \frac{L}{C \left\{ R + j \left(\omega L - \frac{1}{\omega C} \right) \right\}}} \mu v_{g1} \\ &= \frac{\frac{L}{C}}{r_a \left\{ R + j \left(\omega L - \frac{1}{\omega C} \right) \right\} + \frac{L}{C}} \mu v_{g1} \\ &= \frac{\frac{L}{Cr_a}}{R + \frac{L}{Cr_a} + j \left(\omega L - \frac{1}{\omega C} \right)} \mu v_{g1} \quad (2) \end{aligned}$$

CHAPTER XI.—PARA. 23

The scalar value of the instantaneous output voltage is therefore

$$v_a = \frac{\frac{L}{Cr_a}}{\sqrt{\left\{ \left(R + \frac{L}{Cr_a} \right)^2 + \left(\omega L - \frac{1}{\omega C} \right)^2 \right\}}} \cdot \mu v_{g1} \quad (3)$$

and the V.A.F. is $\frac{v_a}{v_{g1}}$ or

$$\text{V.A.F.} = \frac{\frac{L}{Cr_a}}{\sqrt{\left\{ \left(R + \frac{L}{Cr_a} \right)^2 + \left(\omega L - \frac{1}{\omega C} \right)^2 \right\}}} \mu \quad (4)$$

At the resonant frequency $\omega L - \frac{1}{\omega C} = 0$ and the V.A.F. is equal to $\frac{R_d}{r_a + R_d} \mu$. The degree to which other frequencies are attenuated depend upon the ratio $\frac{L}{C}$ and upon the resistances R and r_a . To obtain the highest possible selectivity the circuit constants are so chosen that $\frac{L}{Cr_a} = R$ and under these conditions the V.A.F. at the selected frequency is equal to $\frac{\mu}{2}$. It

should be appreciated that it is theoretically possible to achieve a considerably greater degree of note selectivity than can be usefully employed for C.W. reception, the difficulty of maintaining a heterodyne note of constant frequency being a limiting factor. The curves in fig. 15 illustrate

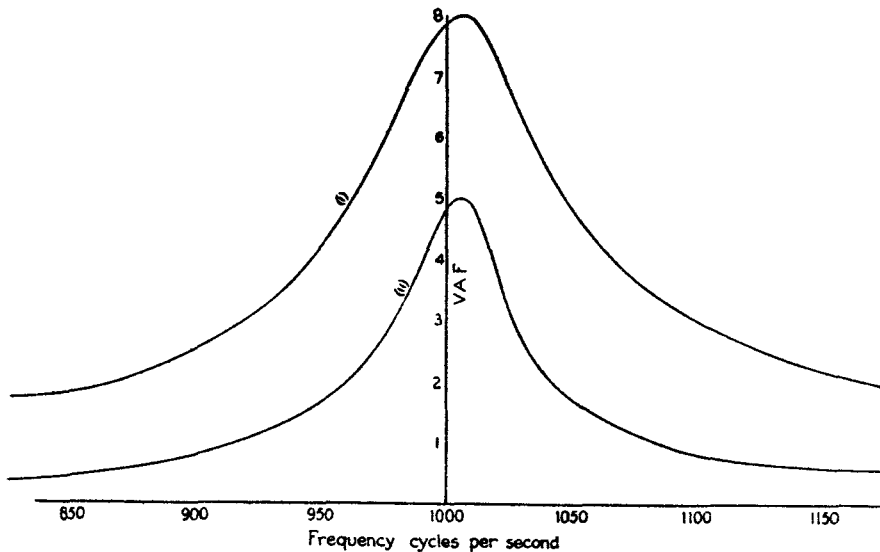


FIG. 15, CHAP. XI.—Response curves of note selectors.

this. Curve (i) is the response curve of a note selector in which $L = .125$ henry, $C = .2\mu F$ and $R = 12.5$ ohms, the valve having an anode A.C. resistance of 12,500 ohms and an amplification factor of 10. Experience shows that suitable chokes for this purpose may be expected to have a resistance of about 100 ohms per henry, i.e. a magnification of 62.5 at 800 cycles per second, the latter frequency being usually adopted for purposes of standardization. Assuming that this

magnification is obtainable at about 1,000 cycles per second, curve (ii) gives the theoretical optimum selectivity obtainable with the same valve. In this case $L = .03125$ henries, $C = .8\mu F$

$R = 3.125$ ohms. It will be seen that the V.A.F. at the selected frequency is now $\frac{\mu}{2}$, but that non-resonant frequencies are greatly attenuated.

24. For the purpose of estimating the variation of signal strength with a given input, it is convenient to draw these response curves on the basis of "decibels below the standard response" against "cycles per second off resonance", using the relation

$$\left. \begin{array}{l} \text{decibels below the standard} \\ \text{response, at frequency } f \end{array} \right\} = 20 \log_{10} \left\{ \frac{\text{V.A.F. at resonant frequency}}{\text{V.A.F. at frequency } f} \right\}$$

Thus, taking curve (i), at the resonant frequency, the V.A.F. is 8, and at 1,025 cycles per second—20 cycles per second off resonance—the V.A.F. is 6.75. $\log_{10} \frac{8}{6.75} = 0.07335$, and the response is $20 \times .07335 = 1.476$ db. below that at resonance. Taking a number of points on curve (i) of fig. 15 in this way, we derive curve (i) of fig. 16. We see that, on the assumption that one signal strength on the arbitrary "audibility scale" corresponds to 6 db., a variation

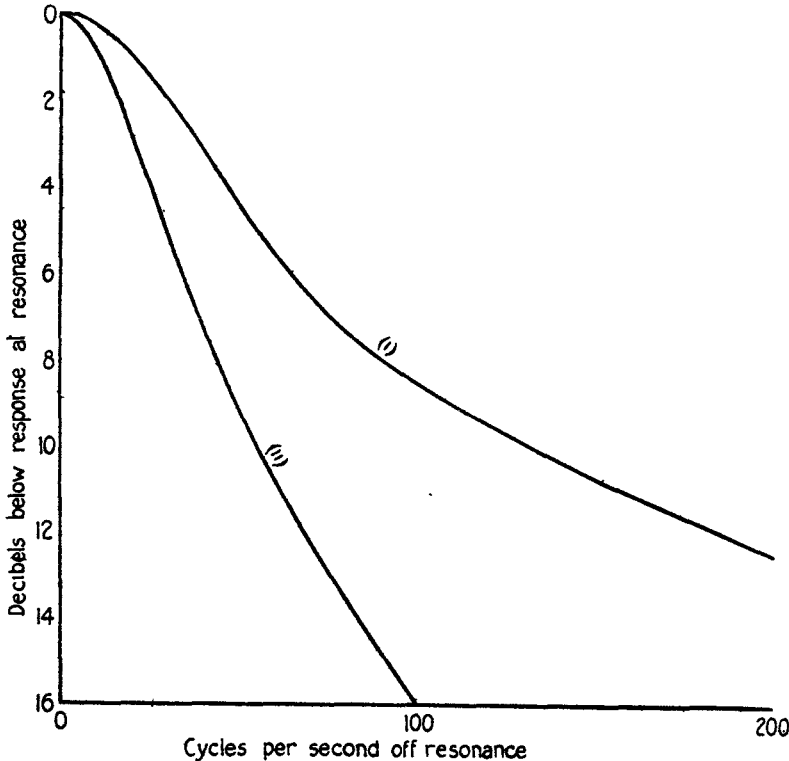


FIG. 16, CHAP. XI.—Curves of FIG. 15 plotted in decibels.

of 200 cycles per second in the heterodyne beat note will cause a variation of two units of signal strength. For this note selector to be of use to the receiving operator, the transmitter frequency must be controlled to within about 200 cycles per second, and preferably to within only 100 cycles per second. This necessitates a stability of 1 part in 10,000 in an operating radio-frequency of 1 Mc/s, even if the heterodyne oscillator is perfectly stable. Curve (ii) of fig. 16, which gives

CHAPTER XI.—PARAS. 25-27

the response in db. corresponding to curve (ii) of fig. 15, shows that a fall of 6 db. will result if the frequency varies only about 30 cycles per second. This degree of stability can only be maintained at comparatively low radio frequencies.

25. To summarize the above, then, it may be said that curves (i) in figs. 15 and 16 respectively show the performance of the best note selector which is at present practicable for general purposes. Where it is possible to employ transmitters of very high frequency stability, i.e. in high power ground-to-ground communication on a single fixed frequency, note selectors of considerably better performance are theoretically possible. In super-heterodyne C.W. receivers (*see paras. 81 et seq.*) the selectivity of the intermediate frequency amplifier or amplifiers must be taken into account and it is found that a note selector having the response of curve (i) is entirely adequate.

Transformer coupling

26. In the transformer-coupled amplifier the anode circuit impedance is the primary winding of a transformer, the secondary voltage being applied to the grid and filament of the succeeding valve, as shown in fig. 17a. This type of amplifier has an advantage over the two previous types, in that the transformer itself may be designed to give a voltage step-up, while no grid condenser and leak is necessary. These advantages are offset to some extent by the greater weight and cost, while even an approach to uniform amplification over a wide frequency band can only be achieved by very careful design of the transformer, and by the choice of a suitable valve for use with it. An exact analysis of the transformer coupled audio-frequency amplifier, for the purposes of obtaining a general expression for the stage gain, is extremely laborious, and even when obtained the expression is difficult to interpret, owing to the large number of circuit constants which must be taken into account. It is however possible to derive comparatively simple expressions for the stage gain at low, medium and high audio frequencies respectively; these are generally only in error by a few parts in one hundred.

27. Referring to fig. 17b, the equivalent circuit is seen to consist of an ideal transformer taking no magnetizing current and having no losses, so that it serves merely to increase the voltage v_a to $Tv_a = v_{g2}$. The magnetizing current is assumed to flow through the inductance L_p which is equal in magnitude to the actual primary inductance. Strictly, the resistance of the primary winding should also be included, in series with L_p . For the present purpose, however, both iron and copper losses may be represented by the resistance R_1 , while $\frac{R_1}{T^2}$ represents the transferred input resistance of the succeeding valve (T_2 of fig. 17(a)). The joint resistance of R_1 and $\frac{R_1}{T^2}$ may be denoted by R_e . Similarly C_e represents the whole of the effective capacitance between anode and filament of T_1 . It is made up of several components, i.e. C_{af} , the inter-electrode capacitance of the valve T_1 , including any distributed capacitance in parallel therewith, T^2C_s , where C_s is the distributed capacitance of the secondary winding plus the input capacitance of T_2 , and $(T-1)^2C_m$ where C_m is the distributed capacitance between the primary and secondary windings. Thus

$$C_e = C_{af} + T^2C_s + (T-1)^2C_m$$

The inductance L_e represents the total leakage inductance of the transformer, transferred to the primary winding. At low audio frequencies, the reactance of C_e is very large. If it is considered to be infinite, the voltage v_a across the primary winding depends only upon the relative magnitudes of ωL_p and R_e ; if R_e is very much greater than ωL_p it may be entirely neglected, so that

$$\begin{aligned} i_a &= \frac{\mu v_{g1}}{\sqrt{r_a^2 + (\omega L_p)^2}} \\ v_a &= \frac{\omega L_p}{\sqrt{r_a^2 + (\omega L_p)^2}} \mu v_{g1} \\ \frac{v_{g2}}{v_{g1}} &= \frac{Tv_a}{v_{g1}} = \frac{\omega L_p}{\sqrt{r_a^2 + (\omega L_p)^2}} T \mu. \quad (\text{"Formula A."}) \end{aligned}$$

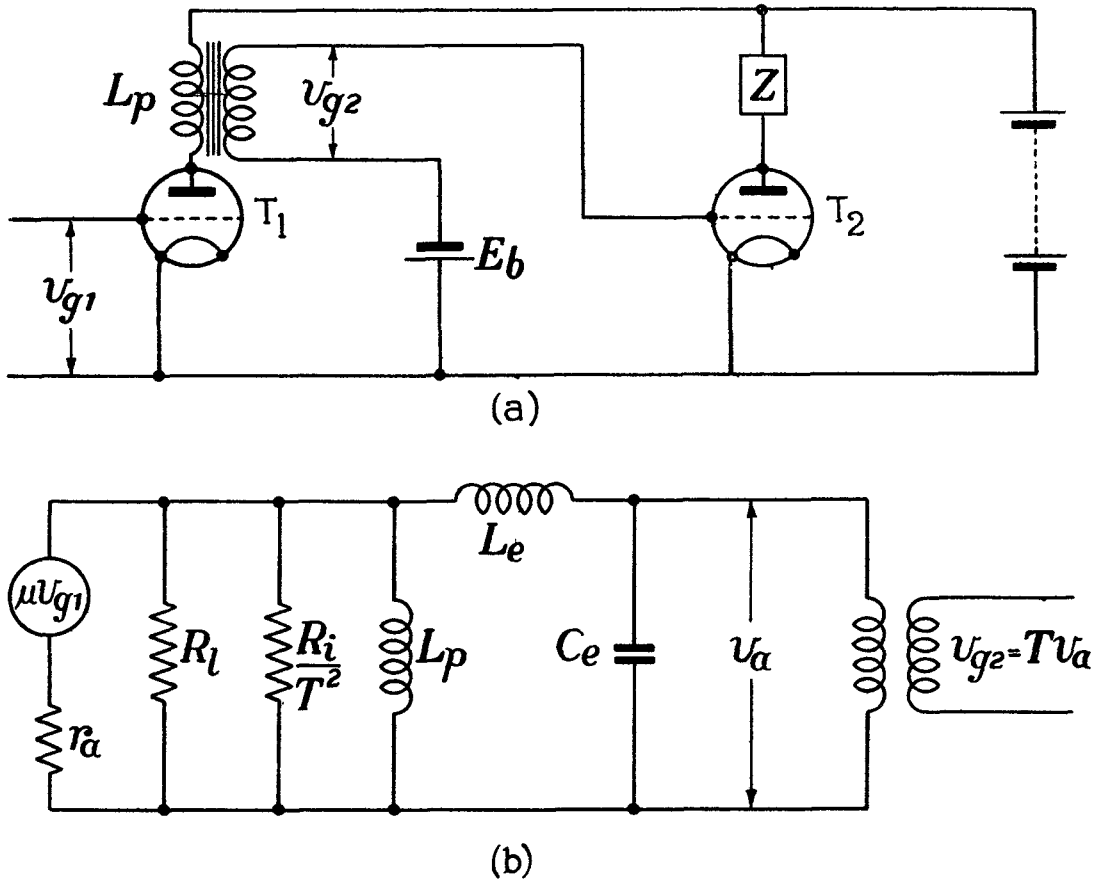


FIG. 17, CHAP. XI.—Transformer-coupled amplifier and equivalent circuit.

At medium audio frequencies, however, while the reactance of C_e is still large compared with R_e , the reactance ωL_p is large compared with R_e also, and the anode load impedance is to all intents and purposes the resistance R_e only. This is of course particularly true at the frequency at which $\omega L_p = \frac{1}{\omega C_e}$ (bearing in mind that L_e has no physical existence but is merely an effective inductance representing magnetic leakage). The stage gain in the region of this frequency will therefore be

$$\frac{v_{g2}}{v_{g1}} = \frac{R_e}{r_a + R_e} T\mu. \quad (\text{"Formula B."})$$

At frequencies appreciably higher than the resonant frequency of L_p , C_e , the current through L_p and R_e will be negligible. The leakage inductance and effective capacitance are then, to all intents and purposes, in series with r_a , and

$$i_a = \frac{\mu v_{g1}}{\sqrt{r_a^2 + \left(\omega L_e - \frac{1}{\omega C_e}\right)^2}}$$

$$\frac{v_{g2}}{v_{g1}} = \frac{\frac{1}{\omega C_e}}{\sqrt{r_a^2 + \left(\omega L_e - \frac{1}{\omega C_e}\right)^2}} T\mu. \quad (\text{"Formula C."})$$

CHAPTER XI.—PARAS. 28-29

28. Before illustrating the above principles by numerical examples, it is necessary to appreciate the magnitude of the quantities C_e , L_e , L_p and R_e , and to discuss the transformation ratio T . At first sight it would appear possible to obtain a very high stage gain merely by increasing T , but this is not so. At low audio frequencies a high V.A.F. can only be obtained by making L_p as large as possible. To obtain a large stage gain by virtue of a high value of T would therefore necessitate a corresponding increase in the number of secondary turns, since

$T = \frac{N_2}{N_1}$. Such an increase would increase the distributed capacitance of the secondary, i.e. would mean an increase in the value of C_s and also in the mutual capacitance C_m . An increase in T therefore increases C_e in greater proportion because $C_e = C_{af} + T^2 C_s + (T - 1)^2 C_m$. Alternatively, if T is increased by a decrease in the primary turns, C_e is still proportional to $T^2 C_s + (T - 1)^2 C_m$ and although C_s and C_m are not increased by a reduction of primary turns, the low-frequency response falls off owing to the reduction of L_p , while the medium-frequency response falls off owing to an increase in iron and copper losses, which is in effect a decrease of R_e . It should be noted that if there were no transformer losses R_e would be infinite and not zero. The transformation ratio therefore lies between limits of from 2 to 10, the higher values being only employed where uniformity of frequency response is unimportant. The capacitance C_e rarely exceeds 3,000 $\mu\mu F$; C_{af} will generally be of the order of 10 $\mu\mu F$, C_s will not usually exceed 200 $\mu\mu F$ and C_m 10 $\mu\mu F$. If $T = 2$,

$$C_e = 10 + 800 + 10 = 820 \mu\mu F,$$

while if $T = 10$

$$C_e = 10 + 2,000 + 810 = 2,820 \mu\mu F.$$

The effective resistance R_e usually lies between 10^5 and 5×10^5 ohms; it depends to some extent upon the transformation ratio and also upon the magnitude of the primary inductance. The effective inductance L_e depends upon the general design of the transformer but is usually about one per cent. of L_p .

29. In the following example the valve is assumed to possess the following constants, viz. $r_a = 10^4$ ohms, $\mu = 10$. The transformation ratio is 3.16 to 1. ($T^2 = 10$) and $R_e = 10^5$ ohms unless otherwise stated while $C_e = 1,500 \mu\mu F$.

Example.—(i) Find the stage gain of a transformer-coupled amplifier at 200 cycles per second if the primary inductance is (a) 10 henries, (b) 50 henries.

$$\begin{aligned} (a) \omega L_p &= 2\pi \times 200 \times 10 \\ &= 12,570 \text{ ohms.} \end{aligned}$$

Since this is small compared with R_e the stage gain is, by formula A,

$$\begin{aligned} &\frac{\omega L_p}{\sqrt{r_a^2 + (\omega L_p)^2}} T \mu \\ &= \frac{1.257 \times 10^4}{\sqrt{(10^4)^2 + (1.257 \times 10^4)^2}} \times 3.16 \times 10 \\ &= \frac{1.257}{\sqrt{1 + 1.257^2}} \times 31.6 \\ &= 24.8. \\ (b) \omega L_p &= 2\pi \times 200 \times 50 \\ &= 62,800 \text{ ohms.} \end{aligned}$$

By formula A,

$$\begin{aligned} \frac{v_{g2}}{v_{g1}} &= \frac{6.28 \times 10^4}{\sqrt{(10^4)^2 + (6.28 \times 10^4)^2}} \times 31.6 \\ &= \frac{6.28}{\sqrt{41}} \times 31.6 \\ &= 30.6. \end{aligned}$$

Since, however, ωL_p is of the same order as R_e , it is possible that formula B may give a more accurate result. By this method

$$\begin{aligned}\frac{v_{g2}}{v_{g1}} &= \frac{R_e}{r_a + R_e} T\mu \\ &= \frac{10^5}{10^4 + 10^5} \times 31.6 \\ &= 28.6.\end{aligned}$$

The latter result is probably more nearly correct than the former.

(ii) If the primary inductance is 50 henries, and the leakage reactance .5 henry, find the stage gain at the frequency for which $\omega L_e = \frac{1}{\omega C_e}$.

By formula C,

$$\begin{aligned}\frac{v_{g2}}{v_{g1}} &= \frac{\frac{1}{\omega C_e}}{\sqrt{r_a^2 + \left(\omega L_e - \frac{1}{\omega C_e}\right)^2}} T\mu \\ &= \frac{1}{\omega C_e r_a} T\mu \\ &= \frac{T\mu}{r_a} \sqrt{\frac{L_e}{C_e}} \\ &= \frac{31.6}{10^4} \sqrt{\frac{.5 \times 10^{12}}{1,500}} \\ &= 57.6,\end{aligned}$$

which is greater than $T\mu$. The frequency at which this occurs, i.e. $\frac{1}{2\pi\sqrt{L_e C_e}}$, is in this particular instance about 5,800 cycles per second.

30. It will be seen that the effect of the subsidiary resonance between L_e and C_e is to give increased amplification at the higher audio frequencies; by careful attention to the relative magnitudes of r_a , L_p , C_e and L_e , the response curve may be made substantially flat up to 8,000 or 10,000 cycles per second. The effect of the magnitude of r_a is somewhat as shown in fig. 18. It will be seen that in order to obtain a high amplification at both ends of the frequency scale, r_a should be as low as possible. Since $\mu = r_a g_m$ and g_m is always made as high as possible, this implies that an even response over a wide frequency range is only obtained by using a valve of low amplification factor. For morse reception, the chief requirement of the transformer is a high turns ratio, giving a high amplification in the neighbourhood of 1,000 cycles per second; reduced amplification of frequencies below about 800 or above 2,000 being an advantage rather than otherwise, for a certain degree of audio-frequency selectivity is then obtained. For reception of R/T, uniform amplification of the band covering from 400 to 2,000 cycles per second will give sufficient intelligibility, but for the reception of entertainment programmes it is usual to aim at even amplification of all frequencies from about 80 to 5,000, and an even wider band is desirable.

CHAPTER XI.—PARA. 30

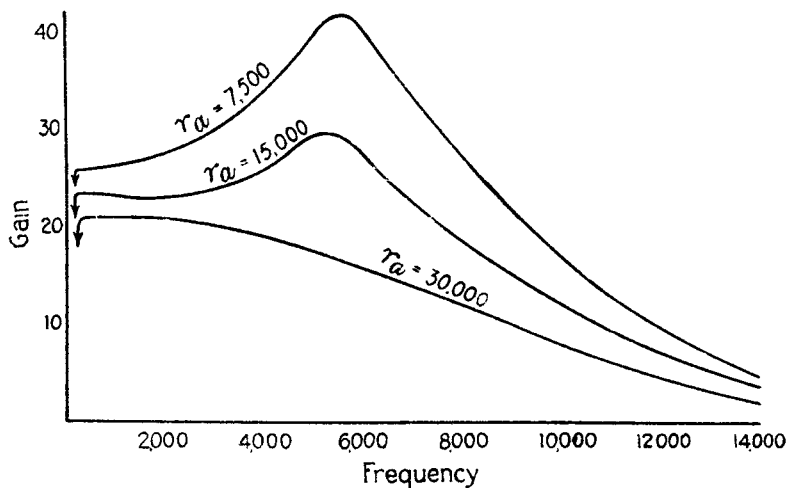


FIG 18, CHAP. XI.—Effect of r_a upon high-note response.

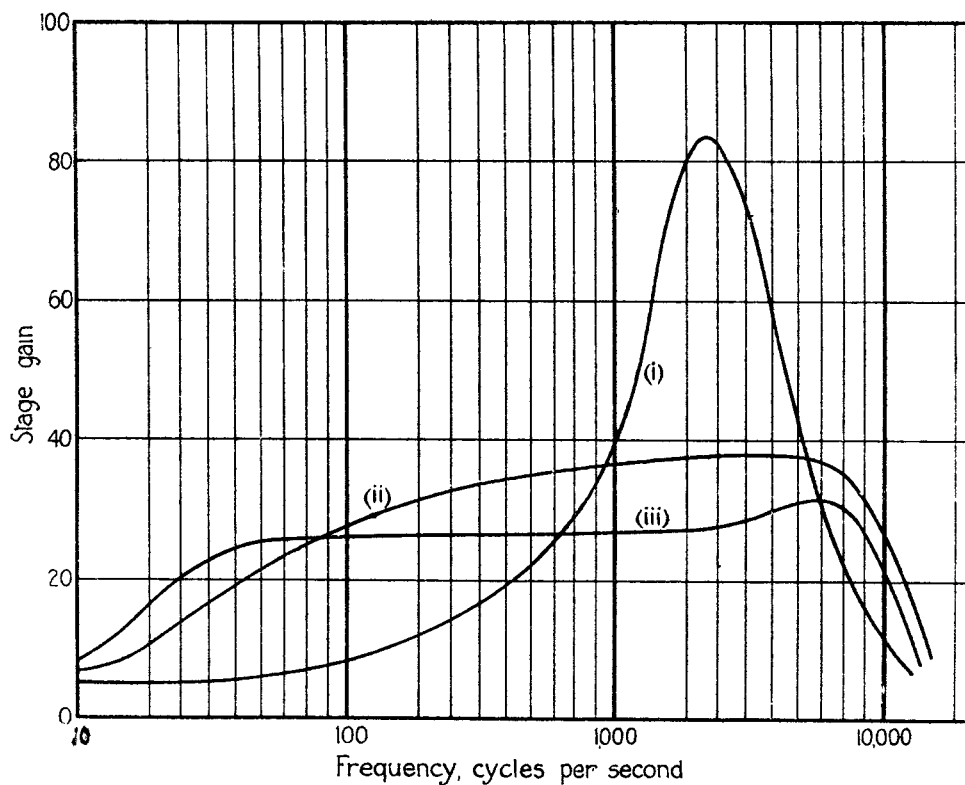


FIG. 19, CHAP. XI.—Response curves of different A.F. transformer

31. These stringent requirements can only be met by designing the transformer with a very high primary inductance and small leakage reactance. For reasons already stated a high primary inductance is only achieved at the expense of a low turns ratio. Fig. 19 shows response characteristics of three different transformers. Curve (i) is that of a transformer perfectly suitable for C.W. reception, giving a high amplification only over the band 1,500 to 4,000 cycles. The primary inductance is 16 henries and the turns ratio 9/1. Curve (ii) is that of a transformer suitable for ordinary R/T communication, the amplification being practically even between 800 and 8,000 cycles. Its primary inductance is 10 henries and turns ratio 6/1. Curve (iii) is that of a transformer designed for reception of broadcast entertainment programmes. Its range of even amplification is practically from 60 to 9,000 cycles per second, the primary inductance being 70 henries and the turns ratio 3·5/1. It should be clearly understood that this comparatively high standard of reproduction is not necessary for the transmission of speech alone, i.e. for commercial or service communication, and may even be a disadvantage when electrical interference exists. The slight increase of gain between 5,000 and 9,000 cycles per second serves to compensate for the high frequency cut-off due to the selectivity of the pre-detector stages.

Parallel-feed transformer coupling

32. The primary inductance of a transformer is reduced by the presence of an appreciable steady flux in the core, such as is caused by the steady component of anode current. In some cases a circuit similar to fig. 20a is adopted in order to avoid this steady magnetization, the steady component of the anode current passing through the feed resistance R_f , and the alternating component through the circuit C, r, L_p . The circuit is then, in effect a combination of resistance and transformer coupling. The capacitance of the blocking condenser C is chosen with regard to the primary inductance, in such a manner that the circuit L_p, C , fig. 20b, is an acceptor for some low audio frequency, e.g. that at which $\omega L_p = r_a$. At this frequency, the opposition offered by the whole external circuit is that of r and R_f in parallel, and the anode current is

$$i_a = \frac{\mu v_{g1}}{r_a + \frac{r R_f}{r + R_f}}$$

Of this current, a fraction $\frac{R_f}{r + R_f}$ flows through the transformer primary, i.e.

$$\begin{aligned} i_L &= \frac{\frac{R_f}{r + R_f}}{r_a + \frac{r R_f}{r + R_f}} \mu v_{g1} \\ &= \frac{R_f}{r_a r + r_a R_f + r R_f} \mu v_{g1} \\ &= \frac{1}{r_a \left(\frac{r}{R_f} + 1 \right) + r} \mu v_{g1}, \end{aligned}$$

and the voltage v_a between the primary terminals is $\omega L_p i_L$ hence

$$\begin{aligned} \text{V.A.F.} \div \frac{v_a}{v_{g1}} &= \frac{\omega L_p}{r_a \left(\frac{r}{R_f} + 1 \right) + r} \mu \\ &\div \frac{\omega L_p}{r_a} \mu, \end{aligned}$$

$$\text{and } \frac{v_{g2}}{v_{g1}} = \frac{\omega L_p}{r_a} T \mu.$$

CHAPTER XI.—PARA. 33

Since $\omega L_p = r_a$, the V.A.F. at the resonant frequency of the circuit $L_p C$ is very nearly equal to μ , and the stage gain equal to $T\mu$, whereas with the usual series connection it is $\frac{T\mu}{\sqrt{2}}$. Provided that the primary inductance L_p is of a high order, e.g. 50 henries, the low-frequency response is greatly improved by the adoption of parallel feed. In order to maintain a high amplification

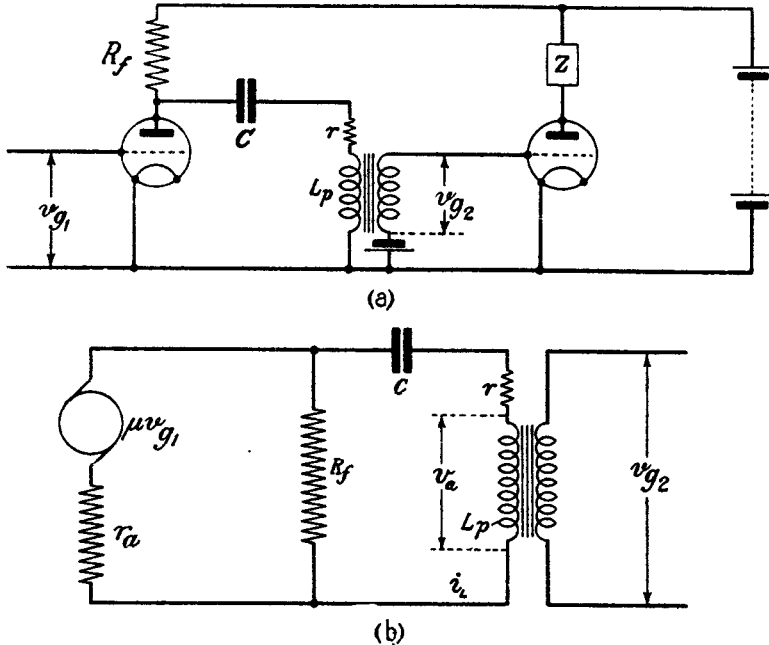


FIG. 20, CHAP. XI.—Parallel feed transformer coupling.

at frequencies above resonance, the feed resistance should be at least equal to $2r_a$. This necessitates an increase in the H.T. supply voltage, and consequently the arrangement is rarely adopted in battery-operated receivers.

Power amplification

33. The anode circuit load impedance in the final stage of an audio-frequency amplifier is the reproducing device, i.e. telephones or loudspeaker. Since the latter can only operate if supplied with electric power the final stage must function as a power amplifier, and may be adjusted to meet either of two requirements. These are (i) maximum power output for a given input voltage, or (ii) maximum undistorted power with the greatest input voltage that can be usefully employed. If it is desired to achieve the first condition, since the maximum power is to be developed irrespective of distortion, grid current may be allowed to flow during positive half-cycles of grid voltage, and the anode current excursion may be allowed to enter the lower curved portion of the characteristic curve of the valve. In either event, the load impedance should preferably be purely resistive, or if this is not practicable, the power factor should be as high as possible.

Conditions for maximum power output

34. The power developed in the output circuit will now be derived, first for maximum power with a given grid voltage. Assuming that the load is a dynamic resistance of R ohms, and the input voltage to be V_g (R.M.S.) the R.M.S. anode current is

$$I_a = \frac{\mu V_g}{r_a + R}$$

The output voltage V_a will be

$$RI_a = \frac{R}{r_a + R} \mu V_g,$$

and the power developed in the load resistance, is

$$P = \frac{R}{(r_a + R)^2} \mu^2 V_g^2.$$

For a given input voltage, the output power is directly proportional to $\frac{R}{(r_a + R)^2}$. This expression is a maximum when $R = r_a$ and the maximum possible output from a given input voltage V_g is

$$P_{\max} = \frac{\mu^2 V_g^2}{4 r_a} = \frac{\mu^2 V_g^2}{8 r_a}$$

where V_g is the peak value of the input voltage. The manner in which the power output varies for different values of the ratio $\frac{R}{r_a}$ is shown in fig. 21. The fall in output resulting from some slight mismatching is not serious, 90 per cent. of the maximum output being obtained when $R = \frac{r_a}{2}$ and also when $R = 2 r_a$.

35. The power obtainable from an amplifier having a reactive load impedance of power factor $\cos \phi$ will depend both upon the magnitude of the impedance and the power factor. This may be demonstrated as follows:—Let the amplifier have an anode load impedance $Z_o = \sqrt{R_o^2 + X_o^2}$, the anode A.C. resistance being denoted by r_a as usual. With an input voltage V_g (R.M.S.) the effective E.M.F. acting in the anode circuit is μV_g , and the resulting anode current I_a . The power dissipated in the load is $I_a^2 R_o = P$. Since

$$\begin{aligned} I_a &= \frac{\mu V_g}{\sqrt{(r_a + R_o)^2 + X_o^2}} \\ P &= \frac{\mu^2 V_g^2 R_o}{(r_a + R_o)^2 + X_o^2} \\ &= \frac{R_o}{r_a^2 + 2r_a R_o + R_o^2 + X_o^2} \mu^2 V_g^2 \\ &= \frac{R_o}{r_a^2 + 2r_a R_o + Z_o^2} \mu^2 V_g^2 \\ &= \frac{\frac{R_o}{Z_o}}{r_a \left(\frac{1}{Z_o} + \frac{2R_o}{Z_o^2} + \frac{1}{r_a} \right)} \mu^2 V_g^2 \end{aligned}$$

CHAPTER XI.—PARA. 35

Since $\frac{R_o}{Z_o}$ is the power factor of the load, i.e. $\cos \varphi$,

$$P = \frac{\cos \varphi}{\frac{r_a}{Z_o} + \frac{Z_o}{r_a} + 2 \cos \varphi} \times \frac{\mu^2 V_g^2}{r_a}$$

For a constant power factor, this will be a maximum when $\frac{r_a}{Z_o} + \frac{Z_o}{r_a}$ is a minimum, or when $r_a = Z_o$. The power output is then

$$\frac{\cos \varphi}{1 + \cos \varphi} \times \frac{\mu^2 V_g^2}{2r_a}$$

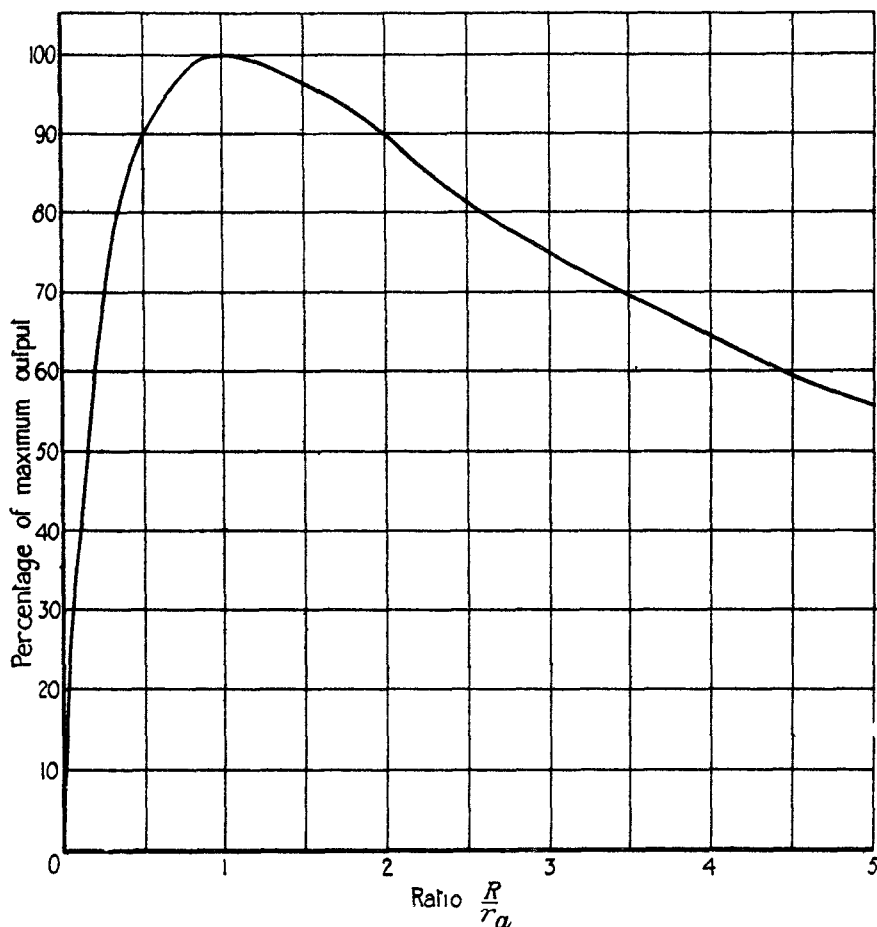


FIG. 21, CHAP. XI.—Variation of power output with ratio $\frac{R}{r_a}$.

For purely resistive loads, $\cos \varphi = 1$, and this expression becomes identical with that derived above.

Example.—Find the power output of a V.R.22 valve when the applied grid-filament voltage has a peak value of 3 volts at 1,000 cycles per second and the anode load is (i) purely resistive, 3,350 ohms, (ii) purely resistive, 13,400 ohms, (iii) an inductance of 1 henry and a resistance of 5,000 ohms, in series, (iv) the power output with optimum resistive load.

The constants of this valve are :— $r_a = 6,700$ ohms, $\mu = 16$. The R.M.S. input is $\frac{3}{\sqrt{2}}$ volts, therefore $V_g^2 = \frac{9}{2} = 4.5$.

In case (i)

$$\begin{aligned}
 P &= \frac{R}{(r_a + R)^2} \mu^2 V_g^2 \\
 P &= \frac{3,350}{(6,700 + 3,350)^2} \times 16^2 \times 4.5 \\
 &= \frac{1}{3,350 \times 99} \times 256 \times 4.5 \\
 &= .0382 \text{ watts} \\
 &= 38.2 \text{ milliwatts.}
 \end{aligned}$$

In case (ii)

$$\begin{aligned}
 P &= \frac{13,400 \times 16^2 \times 4.5}{(6,700 + 13,400)^2} \\
 &= \frac{2 \times 256 \times 4.5}{9 \times 6,700} \\
 &= .0382 \text{ watts or } 38.2 \text{ milliwatts as before.}
 \end{aligned}$$

In case (iii)

$$\begin{aligned}
 Z_o &= \sqrt{R_o^2 + X_o^2} \\
 X_o &= \omega L = 2\pi \times 1,000 \times 1 = 6,280 \text{ ohms} \\
 R_o &= 5,000 \text{ ohms} \\
 Z_o &= \sqrt{40 \times 10^6 + 25 \times 10^6} \\
 &= 1,000\sqrt{65} \\
 &= 8,060 \text{ ohms} \\
 \frac{r_a}{Z_o} &= \frac{6,700}{8,060} = .831 \\
 \frac{Z_o}{r_a} &= 1.2 \\
 \cos \varphi &= \frac{R}{Z_o} = \frac{5,000}{8,060} = .62 \\
 P &= \frac{\cos \varphi}{1 + \cos \varphi} \frac{\mu^2 V_g^2}{2r_a} \\
 &= \frac{.62}{1 + .62} \times \frac{256 \times 4.5}{13,400} \\
 &= .0328 \text{ watts or } 32.8 \text{ milliwatts.}
 \end{aligned}$$

In case (iv)

$$\begin{aligned}
 P &= \frac{\mu^2 V_g^2}{4r_a} \\
 &= \frac{256 \times 4.5}{4 \times 6,700} \\
 &= \frac{18 \times 16}{6,700} \\
 &= .043 \text{ watts or } 43 \text{ milliwatts.}
 \end{aligned}$$

CHAPTER XI.—PARAS. 36-37

36. The operating conditions for maximum power output are illustrated in fig. 22 which gives the $I_a - V_a$ characteristics of a small power valve having an r_a of 5,000 ohms and a g_m of 2.4 milliamperes per volt, so that $\mu = 12$. The power dissipation of this valve is 600 milliwatts. In the valve oscillator, which is only a special form of amplifier, grid excitation is maintained so long as the H.T. supply and grid circuits are completed, so that the D.C. power input may be very much greater than that dissipated by the valve. In receiving amplifiers and certain amplifiers used in transmitters, the power input is maintained even when no grid excitation is applied and consequently the power input must never exceed the permissible dissipation of the valve itself.

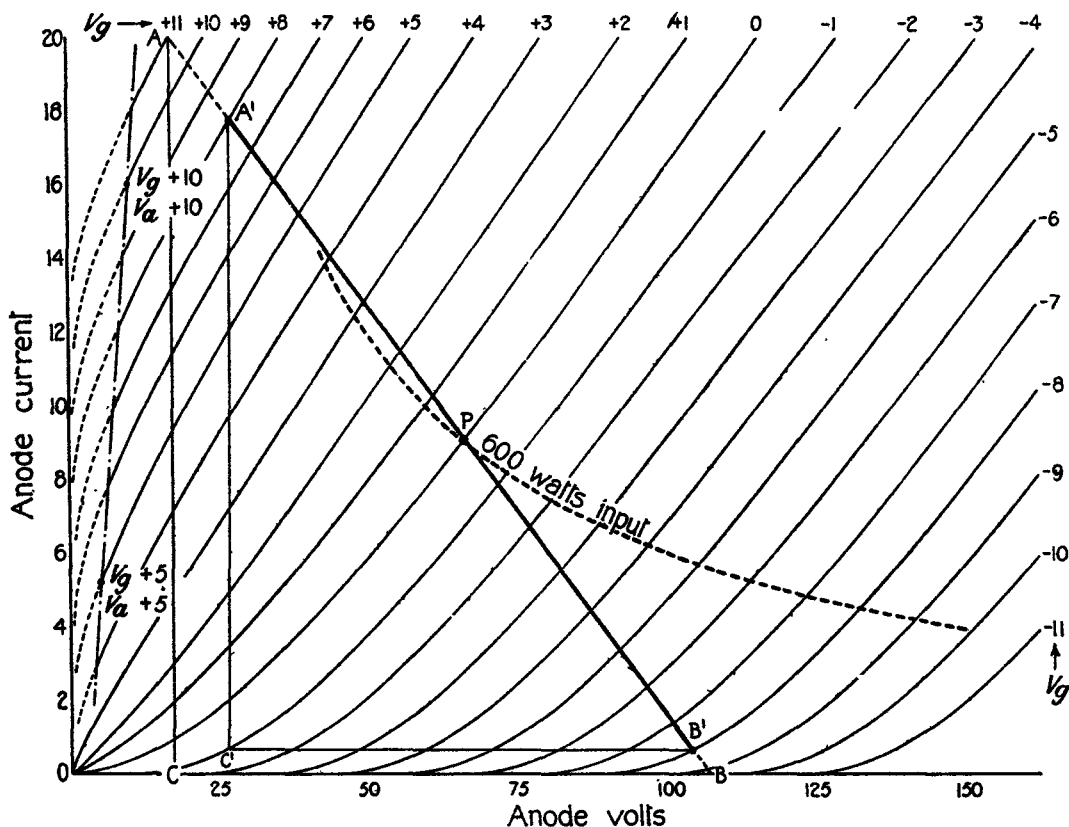


FIG. 22, CHAP. XI.—Operating conditions for maximum power output.

In the present instance then, the operating point P is located at the intersection of the curve $V_g = 0$ and the 600 watts power line. A load line AB corresponding to a resistance of approximately 5,000 ohms has also been drawn through the operating point. The power output is obviously dependent upon the grid swing.

37. In W/T receivers the maximum permissible swing is not subject to limitation in order to avoid distortion but nevertheless is restricted by the fact that the grid must never become positive with respect to the anode, otherwise secondary emission may occur at the grid and the valve will fail to function in the desired manner. The chain-dotted line passing through points such as $(V_g = +5, V_a = +5)$, $(V_g = +10, V_a = +10)$, marks the permissible limit of the positive part of the grid swing due to this factor. With this particular valve and load, it is seen that the peak value of the grid voltage must not exceed about 11 volts, and the anode current will be zero during a portion of the cycle, i.e. about the negative peak of the input grid voltage. It is not possible to develop a generally applicable expression giving the output and

efficiency under these conditions, but as an estimate, it is usual to take the output as equal to one quarter the area of the triangle ACB, i.e. to $\frac{AC \times CB}{8}$. In the figure the area of the rectangle $AC \times CB$ is 20 (milliamperes) $\times$ 90 (volts) or 1,800 milliwatts; the output power is one-eighth of this or 225 milliwatts, and the efficiency $\frac{225}{600} \times 100$ or 37.5 per cent. The output wave-form will of course be considerably distorted.

38. If the input grid swing is limited to say 16 volts so that the load line is A'B' instead of AB, the distortion will be considerably reduced. Provided that the anode current does not reach zero at any point during the cycle, the ratio of second harmonic to fundamental may be estimated by measuring the length of the load line on each side of the operating point. The length of the positive half of the load line being A'P, and the negative half PB', the percentage of second harmonic distortion is

$$\frac{A'P - PB'}{2(A'P + PB')} \times 100$$

Example.—In the original of fig. 22 $A'P = 5\frac{7}{16}$ inches, $PB' = 5\frac{1}{4}$ inches. Find the percentage of second harmonic introduced.

Expressing A'P and PB' in sixteenths of an inch, this gives the percentage of second harmonic as

$$\begin{aligned} & \frac{87 - 42}{2(87 + 42)} \times 100 \\ &= \frac{45}{2 \times 129} \times 100 \\ &= 17.5 \text{ per cent.} \end{aligned}$$

The power output under these conditions is $\frac{A'C' \times B'C'}{8}$, which is rather less than 180 milliwatts, and the efficiency nearly 30 per cent.

Maximum undistorted output

39. (i) If the second of the above conditions is to be fulfilled, grid current must not be allowed to flow during any portion of the cycle, and the anode current excursion must be confined to the practically straight portions of the $I_a - V_g$ curves. As in the previous instance the load impedance is preferably non-reactive. The curves shown in fig. 23 are somewhat idealized, but will serve to illustrate what is desirable. It is seen that the curves are straight lines except at the extreme foot, i.e. below $I_a = 1$ milliampere. To obtain maximum undistorted output, certain relations between the supply voltage, the grid bias and the load resistance must be fulfilled. The optimum load resistance is equal to $2r_a$. The proof of this is exactly as given in Chapter IX for the case of a sinusoidal oscillator operating without grid current.

(ii) Bearing in mind that grid current must be avoided and that the D.C. input must not exceed the permissible dissipation of the valve, the operating point P may be located as follows. Draw the minimum anode current line ST cutting off the curved portion of the characteristics. Erect a convenient voltage ordinate AB where A lies on the curve $V_g = 0$ and B on the line ST. Locate the point C, making $CB = \frac{1}{2} AB$. From S, draw the straight line SC producing it to intersect the dissipation (i.e. input power) line at P. Then P is the desired operating point. The optimum load line is now easily found. Through P draw the current ordinate $P'P''$, intersecting the voltage ordinate AB at P'. Produce the line AB indefinitely in the upward direction, and locate the point D in such a manner that $P'D = 2 P'C$. Draw the line DP, producing it to meet the anode voltage base line at F. The required load line then lies upon DF. The maximum grid swing is defined by the intersection of this line with the characteristic $V_g = 0$ and its intersection with the minimum current line ST.

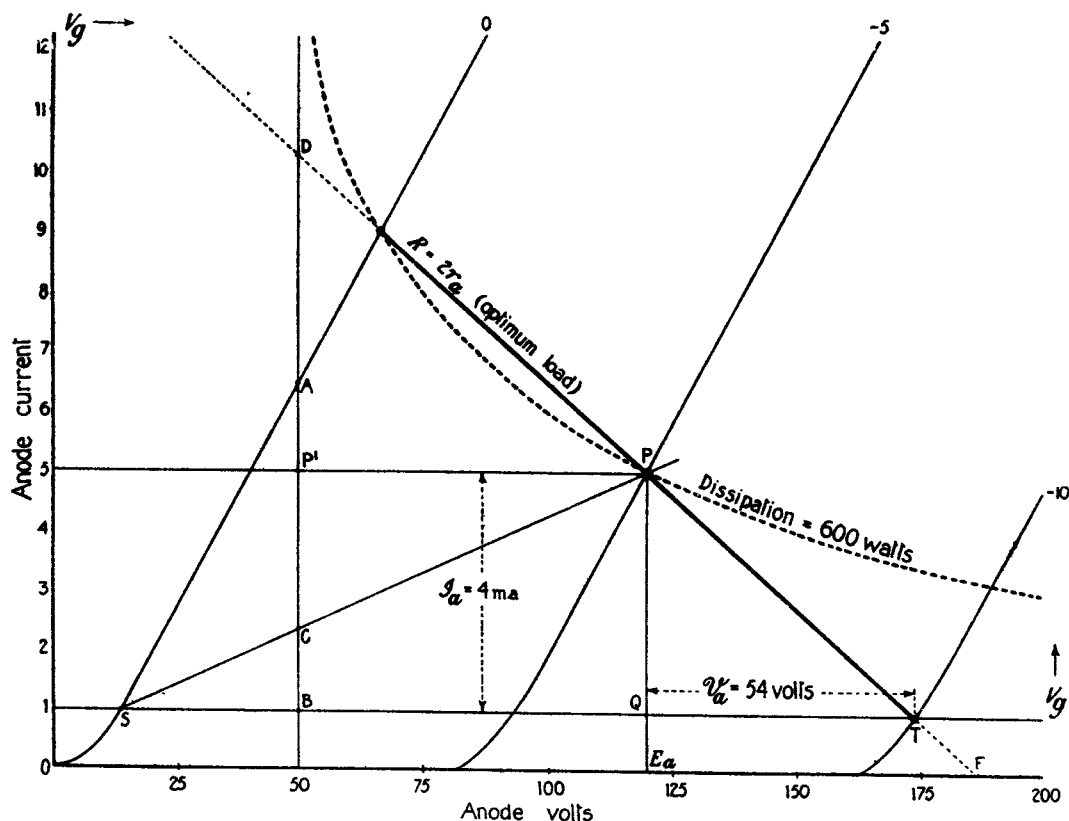


FIG. 23; CHAP. XI.—Theoretical operating conditions, for maximum output without distortion.

40. The co-ordinates of the operating point P give the correct negative grid bias and H.T. voltage for maximum undistorted output, and the peak input grid-filament voltage is equal to the grid bias. In fig. 25 this is 5 volts. The power output is equal to the area of the triangle P Q T. It is seen that $PQ = I_a = 4$ milliamperes, $QT = V_a = 54$ volts. The power output is $\frac{V_a I_a}{2} = \frac{54 \times 4}{2} = 108$ milliwatts. Alternatively, the same result may be derived algebraically from the relation

$$P = \frac{\mu^2 V_g^2}{2} \times \frac{R}{(r_a + R)^2}$$

Since $R = 2r_a$

$$\begin{aligned} P &= \frac{\mu^2 V_g^2}{2} \times \frac{2r_a}{9r_a^2} \\ &= \frac{\mu^2 V_g^2}{9r_a} \end{aligned}$$

$$r_a = 6,500$$

$$g_m = 2.44 \text{ milliamperes per volt}$$

$$\mu = 16$$

CHAPTER XI.—PARAS. 42-43

becomes CD instead of AB, assuming that the full grid swing of 9 volts is to be applied. Before a final decision is made the percentage of distortion must be checked. In the original drawing $CP_1 = 7$ inches, $P_1D = 5$ inches and the percentage of distortion is

$$\frac{7-5}{2(7+5)} \times 100 = \frac{100}{12} = 8\frac{1}{3} \text{ per cent.}$$

The output power is equal to one-eighth of the rectangle CC^1, C^1D . Here $CC^1 = 10$ milliamperes, $C^1D = 84$ volts; the output is 105 milliwatts and the efficiency 17.5 per cent. Now suppose the load resistance to be increased, giving the load line EF ($R = 12,500$ ohms). The output is then only 101 milliwatts and the efficiency about 16 per cent. The distortion is, however, reduced to about 4 per cent.

Figure of merit

42. (i) It has now been shown that the output of an audio-frequency amplifier must always be arrived at as the result of a compromise with regard to the permissible distortion. It is also apparent that to obtain an appreciable power output, a sufficient input grid-filament voltage and a large H.T. supply voltage must be available. Unless these requirements are met it is possible that a power valve may give a smaller output than one not specifically designed for this purpose. In other words, a general purpose triode is often used as an output valve when the input grid swing and H.T. supply voltage are insufficient to obtain increased output by the employment of a power valve.

(ii) It has also been shown that the power output of a given valve is proportional to $\frac{\mu^2}{r_a}$ and the latter is called the figure of merit for any particular valve. The figure of merit of certain valves used in the service is tabulated below.

Valve.	μ	r_a	Figure of merit.
V.R.12F	6	18,000	·002
V.R.19	8.5	5,000	·014
V.R.21	13.7	11,500	·016
V.T.20	8	3,700	·017
V.R.22	16	6,700	·038
V.T.23	5	2,000	·0125

The highest figure of merit in this series is that of the valve V.R.22 which is specifically designed as an output valve handling an input swing of about 9 volts. Valves such as the V.T.20 and V.T.23 will give a greater output, but only if a greatly increased grid swing is available. When it is necessary to provide a power output greater than about 100 milliwatts several alternatives present themselves, such as the use of power triodes in parallel or push-pull, a single pentode valve, or pentodes in push-pull. As these requirements are rarely called for except for R/T reception, they are dealt with in Chapter XII.

Use of output transformer

43. In most cases it is not convenient to design the output impedance especially to match one particular type of valve; the ordinary telephone receiver is, for example, used in conjunction with several of the valves specified above in different service receivers. Fig. 21 shows that the ratio $\frac{R}{r_a}$ may vary from .5 to 2 without causing an appreciable change in the power output, consequently, when the magnitude of the load impedance differs only slightly from the anode A.C. resistance of the valve, the load may be connected in series as has hitherto been assumed. Where the load impedance differs greatly from the anode A.C. resistance of the valve it is

necessary to use an output transformer. The connections are then as in fig. 25a, the primary winding being connected in series in the anode circuit and the load impedance to the secondary winding. If this load is Z ohms, and the transformation ratio of the transformer is T , the effective load transferred to the primary is $\frac{Z}{T^2}$ ohms, the power factor being practically unaltered.

Hence, if the transferred primary load is to equal the anode A.C. resistance of the valve,

$$T = \sqrt{\frac{Z}{r_a}},$$

while for maximum undistorted output,

$$T = \sqrt{\frac{Z}{2r_a}}$$

44. When the output valve is of the small power type, e.g. the valve receiving V.R.22, or of course any type requiring a greater D.C. power supply, it is however inadvisable to connect the telephones or loudspeakers directly in circuit for two reasons. First, the comparatively

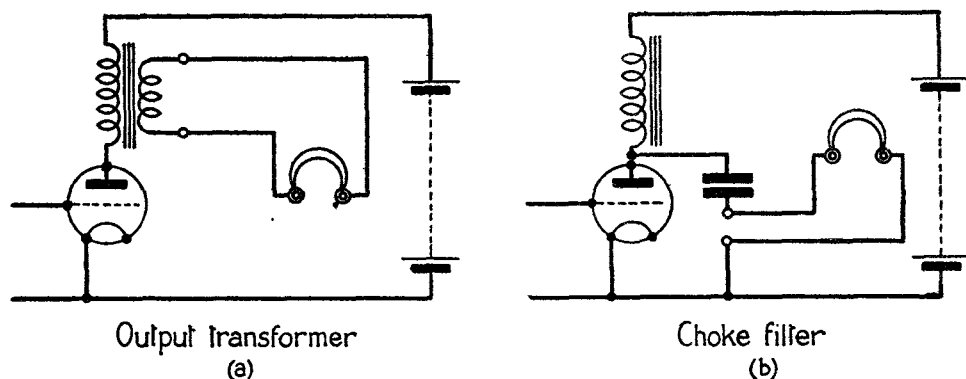


FIG. 25, CHAP. XI.—Typical output circuits for A.F. power amplifier.

large steady component of anode current may lead to overheating of the windings and subsequent breakdown; second, the steady flux caused by this current may be sufficient to saturate or depolarize the magnetic circuit, leading to an inferior response; third, as it is customary to insert the telephones in circuit by means of a plug and jack, the withdrawal of the telephone plug will cause a heavy induced E.M.F. in the anode circuit, which may lead to damage to valves or other components. Even if it is not necessary to use a step-up or step-down transformer for matching purposes, it is usual to use a 1/1 output transformer, which is usually an auto-transformer and is spoken of as the "output choke." As shown in fig. 25b, the telephones are connected across the ends of this inductance, a condenser of large capacitance being inserted in series with them, so that the steady anode current is carried by the choke only. Alternatively the choke and condenser may be said to form a filter, allowing direct current to pass through the choke but not through the telephones, and audio-frequency current to pass through the telephones, but only to a limited extent through the choke. For this filtering action to be efficient, the inductance of the choke should be of the order of ten times that of the telephones.

Radio-frequency amplification

45. The forms of intervalve coupling in general use for radio-frequency amplification are :—

- (i) Tuned-anode capacitance coupling.
- (ii) Tuned transformer coupling.

CHAPTER XI.—PARAS. 46-47

Resistance-capacitance coupling and a semi-a-periodic form of transformer coupling were formerly employed to some extent for amplification of frequencies below about 500 kc/s. The advantage of these, namely, fairly uniform amplification of a wide frequency band without the necessity for any tuning adjustment, is more than offset by their lack of selectivity.

Tuned-anode capacitance coupling

46. Fig. 26 shows a tuned aerial circuit, $L_a C_a$ in which a signal voltage developed across the inductance L_a is applied to the grid and filament of the triode T_1 . The anode circuit consists of an inductance L and capacitance C in parallel, the inevitable losses in this circuit being represented by the resistance R . The input voltage v_{g1} may be of the order of only a few millivolts. It has been shown earlier in this chapter that for signals of the frequency to which the anode circuit is tuned, the V.A.F. of the triode and its tuned anode circuit is

$$\frac{v_a}{v_{g1}} = \frac{R_d}{r_a + R_d} \mu$$

$$\text{where } R_d = \frac{L}{CR}.$$

The grid condenser C_g serves to apply this voltage to the grid of the succeeding valve T_2 , the requirements being that its reactance at the operating frequency shall be small compared with

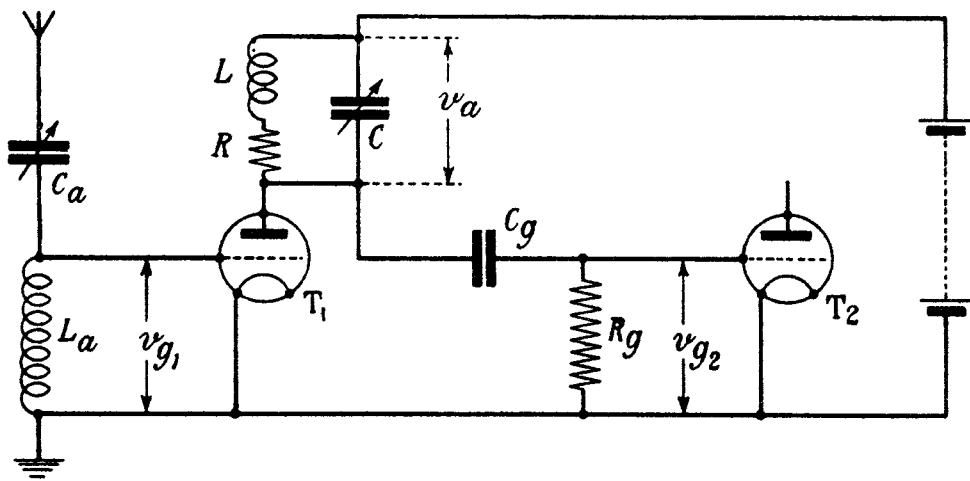


FIG. 26, CHAP. XI.—Tuned anode R.F. amplifier.

the impedance of the grid-filament path, and that its insulation resistance shall be extremely high. Both these conditions are easily fulfilled by a mica dielectric condenser having a capacitance of the order of $\cdot 0001 \mu F$. The grid leak R_g may be from $\cdot 1$ to 4 megohms depending in part upon the function of the valve T_2 , which may be required to act as an additional R.F. amplifier or as a detector valve. The introduction of the grid leak will cause the stage gain to be somewhat less than the V.A.F. as calculated from the above expression, for its resistance must be considered to be in parallel with the dynamic resistance R_d . As however the latter is usually much less than the grid leak resistance, the effect is of minor importance. The input impedance of the following valve must also be taken into consideration, as will be shown later.

47. At frequencies other than the resonant frequency of the anode circuit, the V.A.F., and therefore the stage gain, is less than at the resonant frequency. Provided that the resistance

of the tuned anode circuit is small compared with the ratio $\frac{L}{C}$, the V.A.F. at any frequency $\frac{\omega}{2\pi}$ is given by the equation :—

$$\text{V.A.F.} = \frac{\frac{L}{Cr_a}}{\sqrt{\left(\frac{L}{Cr_a} + R\right)^2 + \left(\omega L - \frac{1}{\omega C}\right)^2}} \mu.$$

(This expression is derived in paragraph 23).

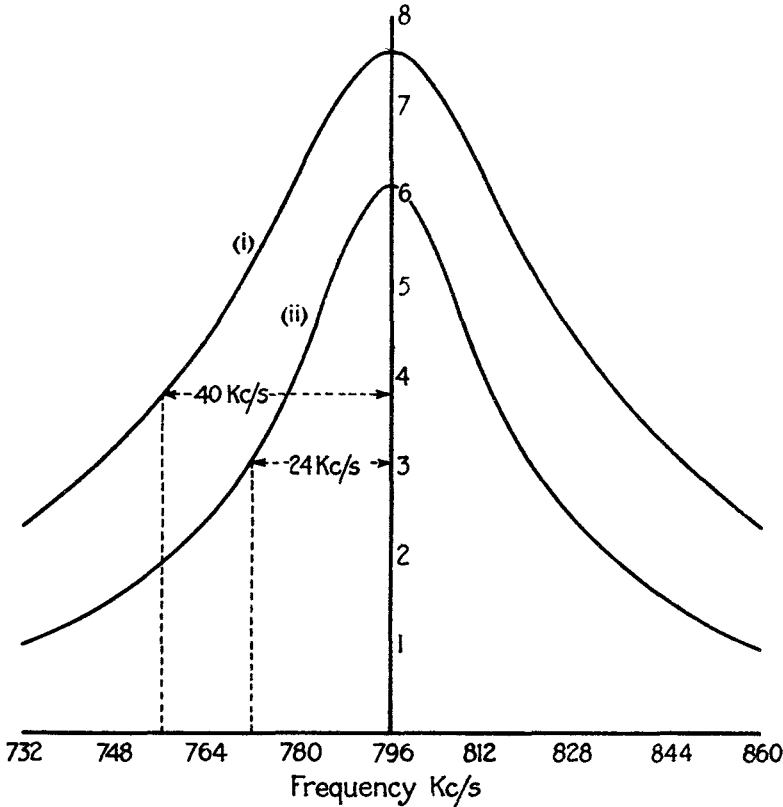


FIG. 27, CHAP. XI.—Response curves with different $\frac{L}{C}$ ratios.

At the resonant frequency of the circuit, $\omega L = \frac{1}{\omega C}$, and this expression simplifies to that already given. The tuned anode amplifier therefore possesses the property of selectivity, and the degree of discrimination is dependent chiefly upon the ratio $\frac{L}{C}$. A large value of $\frac{L}{C}$ gives high amplification at the resonant frequency, but the selectivity is poor, while a low ratio $\frac{L}{C}$ gives less amplification but greater selectivity. The stage gain of a certain tuned-anode capacitance-coupled amplifier in the region of the resonant frequency, 796 kc/s, is shown in fig. 27. The

CHAPTER XI.—PARAS. 48–49

valve used has the following constants, viz. $r_a = 20,000$ ohms, $\mu = 10$. Curve (i) shows the V.A.F. when the inductance L is $160 \mu H$, the capacitance $0.00025 \mu F$, and the resistance 10 ohms, while curve (ii) shows the V.A.F. obtained by halving the inductance and doubling the capacitance the resistance being also reduced to 5 ohms. It will be observed that the magnification of the circuit is the same in each case, i.e. 80 , but the selectivity of the second arrangement is greater than that of the first; for example, with the circuit of large $\frac{L}{C}$ ratio the amplification of a signal 40 kc/s off resonance is one-half that of the desired signal, whereas with the circuit of lower $\frac{L}{C}$ ratio the same reduction is obtained if the undesired signal is only 24 kc/s off resonance. In Chapter V it is shown that the greatest selectivity is achieved when the dynamic resistance of the anode circuit is equal to the anode A.C. resistance of the valve, and the V.A.F. is then equal to $\frac{\mu}{2}$.

Tuned transformer coupling

48. This is an alternative kind of R.F. inter-valve coupling, which may take either of three forms, namely (a) with primary winding tuned to the frequency of the desired signal, (b) with the secondary winding tuned, or (c) with both windings tuned (fig. 28). The latter form is rarely,

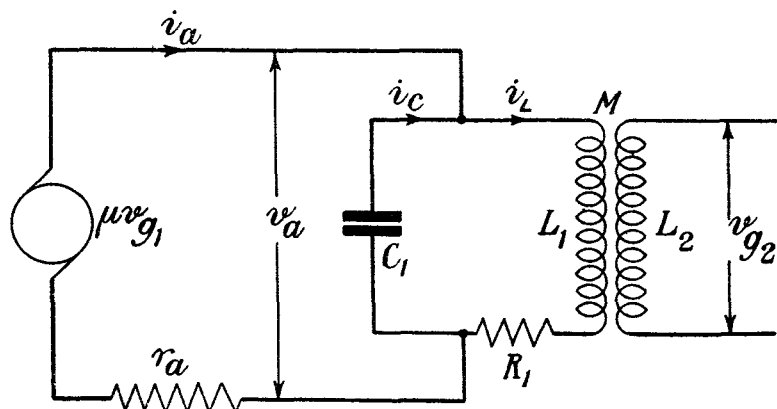
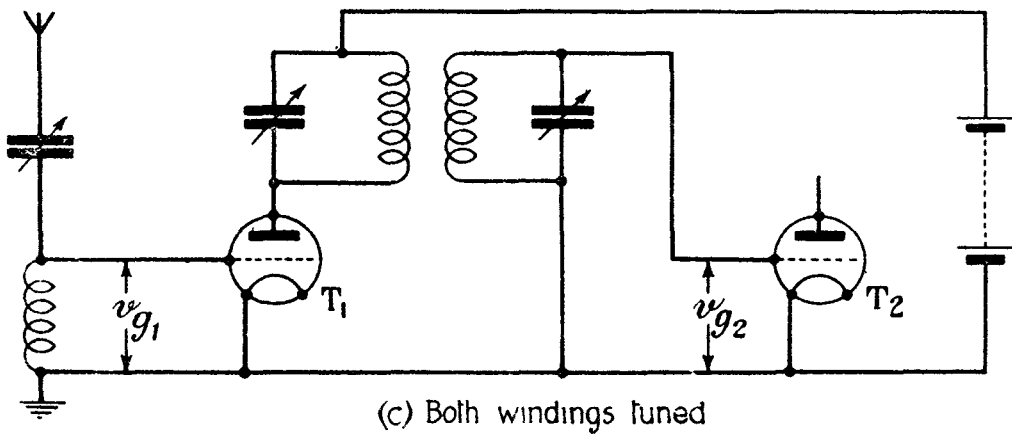
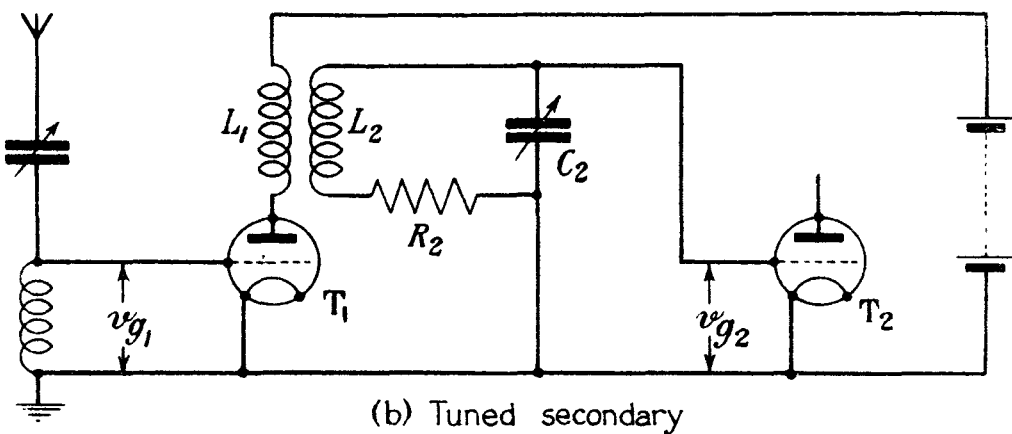
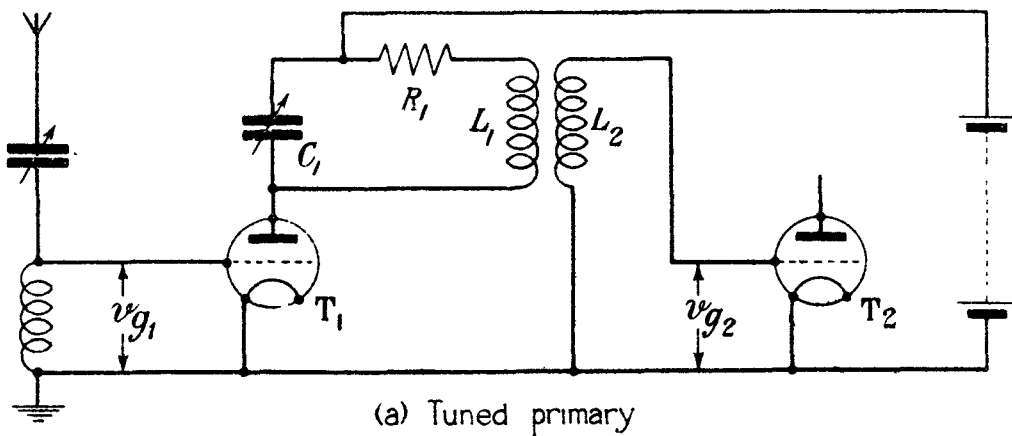


FIG. 29, CHAP. XI.—Equivalent circuit, "tuned primary" coupling.

if ever, adopted when it is necessary to vary the frequency to which the circuits are adjusted, but may be adopted in amplifying stages designed to operate at a single fixed frequency, such as the intermediate frequency stages of a supersonic heterodyne receiver (*see paras. 81 et seq.*). When an R.F. amplifying stage is required to be adjustable to any frequency within a fairly wide range, by means of a variable condenser or inductance, the choice lies between the first and second named arrangements. The stage gain and selectivity of the current shown in fig. 28b are superior to that of fig. 28a, and the former is almost universally employed. The relative selectivity and gain of the two arrangements may now receive a brief consideration.

"Tuned primary" circuit

49. The stage gain of this arrangement is easily derived, for reference to the equivalent circuit of the amplifier (fig. 29) shows that, provided the secondary winding is on open circuit,



TYPES OF R.F. TRANSFORMER COUPLING

the voltage across the primary winding of the coupling transformer is, by analogy with the tuned anode circuit of fig. 26,

$$v_a = \frac{\frac{L_1}{C_1 r_a}}{\sqrt{\left(R_1 + \frac{L_1}{C_1 r_a}\right)^2 + \left(\omega L_1 - \frac{1}{\omega C_1}\right)^2}} \mu v_{g1}.$$

The primary current i_L will be very nearly $\frac{v_a}{\omega L_1}$ and the E.M.F. induced in the secondary winding is $\omega M i_L$ or $\frac{M}{L_1} v_a$. As the secondary is unloaded this is also the secondary terminal P.D., v_{g2} , hence

$$\frac{v_{g2}}{v_{g1}} = \frac{\frac{M}{C_1 r_a}}{\sqrt{\left(R_1 + \frac{L_1}{C_1 r_a}\right)^2 + \left(\omega L_1 - \frac{1}{\omega C_1}\right)^2}} \mu.$$

At the resonant frequency of the primary circuit $\omega L_1 = \frac{1}{\omega C_1}$ and

$$\frac{v_{g2}}{v_{g1}} = \frac{\frac{M}{C_1 r_a}}{R_1 + \frac{L_1}{C_1 r_a}} \mu.$$

If k is the coefficient of coupling between the windings of the transformer, $M = k\sqrt{L_1 L_2}$ and

$$\frac{v_{g2}}{v_{g1}} = \frac{\frac{k\sqrt{L_1 L_2}}{C_1 r_a}}{R_1 + \frac{L_1}{C_1 r_a}} \mu.$$

To simplify still further, multiply the numerator of the right-hand member by $\frac{L_1}{L_1} = 1$. It then becomes

$$\frac{k L_1 \sqrt{L_1 L_2}}{L_1 C_1 r_a} = k \sqrt{\frac{L_2}{L_1}} \times \frac{L_1}{C_1 r_a}.$$

Denoting $\sqrt{\frac{L_2}{L_1}}$ by T , and $\frac{L_1}{C_1 R_1}$, which is the dynamic resistance of the tuned primary circuit, by R_d , we have as a final expression, for the stage gain at the resonant frequency,

$$\frac{v_{g2}}{v_{g1}} = \frac{R_d}{r_a + R_d} k T \mu.$$

50. If the primary and secondary windings are of the same size and shape, for example, if they are wound side by side on a slotted former and the wire gauges so chosen that they occupy equal volumes of winding space, T is equal to the ratio $\frac{\text{Secondary turns}}{\text{Primary turns}}$. The product kT is the effective transformation ratio of the transformer and is less than T because the coefficient of coupling must be less than unity. It may appear possible to increase the stage gain without limit by increasing the ratio $\frac{L_2}{L_1}$, but this is not so, because such an increase must result in a corresponding reduction of k owing to the increased flux leakage between the windings. An

CHAPTER XI.—PARA. 51

increase in the number of secondary turns will also lead to an increase in the distributed capacitance of the winding, and if the resonant frequency of the secondary winding approaches that of the primary, the stage gain is not given by the expression developed above.

"Tuned secondary" circuit

51. Turning now to the circuit given in fig. 28b and using the notation given in the equivalent circuit diagram (fig. 30), the stage gain is found to be

$$\frac{v_{g2}}{v_{g1}} = \frac{\left(\frac{k}{T}\right)^2 \frac{L_2}{C_2 r_a}}{\sqrt{\left[R_2 + \left(\frac{k}{T}\right)^2 \frac{L_2}{C_2 r_a}\right]^2 + \left(\omega L_2 - \frac{1}{\omega C_2}\right)^2}} \frac{T}{k} \mu$$

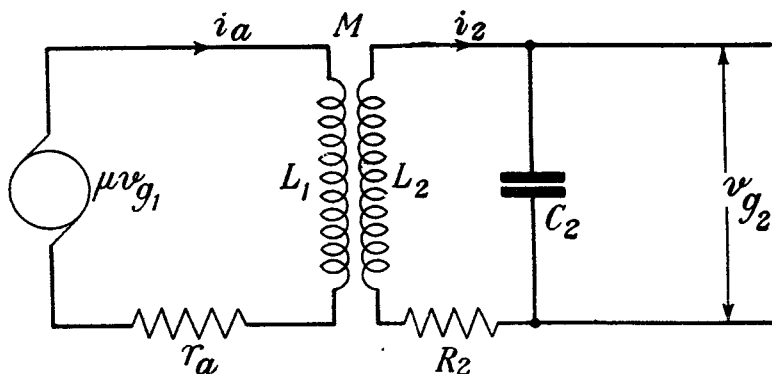


FIG. 30, CHAP. XI.—Equivalent circuit, "tuned secondary" coupling.

where k and T are as previously defined. At the resonant frequency $\omega L_2 - \frac{1}{\omega C_2} = 0$ and

$$\frac{v_{g2}}{v_{g1}} = \frac{\left(\frac{k}{T}\right)^2 R_d}{r_a + \left(\frac{k}{T}\right)^2 R_d} \frac{T}{k} \mu$$

where $R_d = \frac{L_2}{C_2 R_2}$. The effective transformation ratio is now $\frac{T}{k}$, whereas in the previous arrangement it was kT . This is because in the "tuned primary" circuit, the chief losses are due to the comparatively large circulating current in the primary. As the secondary circuit is assumed to be loss-free, an increase in the coupling coefficient k gives a larger secondary E.M.F. and terminal P.D. When the secondary winding is tuned, however, it becomes the seat of a circulating current and its losses are transferred to the primary circuit. The tighter the coupling, the larger is the equivalent primary resistance, and the smaller is the secondary E.M.F. and the P.D. at the condenser terminals. Maximum stage gain and selectivity are obtained when the

transferred resistance $\left(\frac{k}{T}\right)^2 R_d$ is equal to r_a . For a given valve and secondary circuit, the

optimum transformation ratio is found to be equal to $\sqrt{\frac{R_d}{r_a}}$ and the stage gain is then $\frac{\mu}{2} \sqrt{\frac{R_d}{r_a}}$.

Example.—In fig. 30 let $L_2 = 160 \mu H$, C_2 be variable from $.00005 \mu F$ to $.0005 \mu F$, $r_a = 20,000$ ohms, $\mu = 10$. Assuming that R_d is equal to 10 ohms over the whole tuning range and that a coupling coefficient of .9 can be achieved, find the primary inductance for optimum gain at 796 kc/s ($C_2 = .00025 \mu F$), and the gain when $C_2 = .00005 \mu F$, $.00025 \mu F$ and $.0005 \mu F$.

$$\text{When } C_2 = .00025 \mu F, R_d = \frac{L_2}{C_2 R_2} = 64,000 \text{ ohms.}$$

$$\text{The optimum value of } \frac{T}{k} \text{ is } \sqrt{\frac{R_d}{r_a}} = \sqrt{\frac{64,000}{20,000}} = 1.79$$

$$T = 1.79 k = \sqrt{\frac{L_2}{L_1}}$$

$$\begin{aligned} \frac{L_2}{L_1} &= 1.79^2 k^2 \\ &= 3.2 \times .81 \\ &= 2.6 \end{aligned}$$

$$\begin{aligned} L_1 &= \frac{L_2}{2.6} \\ &= 61.5 \mu H. \end{aligned}$$

The gain at this frequency will be $\frac{\mu}{2} \sqrt{\frac{R_d}{r_a}}$ or

$$\begin{aligned} \frac{v_{g2}}{v_{g1}} &= \frac{10}{2} \times 1.79 \\ &= 8.95. \end{aligned}$$

When $C_2 = .00005 \mu F$, $R_d = 320,000$ ohms.

$$\frac{v_{g2}}{v_{g1}} = \frac{\left(\frac{k}{T}\right)^2 R_d}{r_a + \left(\frac{k}{T}\right)^2 R_d} \frac{T}{k} \mu$$

$$\frac{T}{k} = 1.79$$

$$\frac{k}{T} = .558$$

$$\left(\frac{k}{T}\right)^2 = .312$$

$$\begin{aligned} \left(\frac{k}{T}\right)^2 R_d &= .312 \times 320,000 \\ &= 100,000 \text{ ohms.} \end{aligned}$$

$$\begin{aligned} \frac{v_{g2}}{v_{g1}} &= \frac{100,000}{100,000 + 20,000} \times 1.79 \times 10 \\ &= 14.9, \end{aligned}$$

and when $C_2 = .0005 \mu F$, $R_d = 32,000$ ohms

$$\begin{aligned} \left(\frac{k}{T}\right)^2 R_d &= .312 \times 32,000 \\ &= 10,000 \text{ ohms} \end{aligned}$$

$$\begin{aligned} \frac{v_{g2}}{v_{g1}} &= \frac{10,000}{10,000 + 20,000} \times 10 \times 1.79 \\ &= 5.97 \end{aligned}$$

CHAPTER XI.—PARAS. 52-53

52. For purposes of comparison, fig. 31a gives the respective response curves (stage gain against kilocycles per second off resonance) of (i) the transformer-coupled amplifier with tuned primary, and (ii) the same amplifier with tuned secondary. The resonant frequency is 796 kc/s, the valve has an r_a of 20,000 ohms and a μ of 10, while the constants of the tuned circuit are the same in each case, namely $L = 160 \mu H$, $C = .00025 \mu F$, $R = 10$ ohms. In fig. 31b, the relative gains are plotted in decibels with reference to an arbitrary level corresponding to a stage gain of 8. It is seen that at the resonant frequency the tuned secondary arrangement is the better by about 1.5 db. The tuned secondary attenuates a signal 34 kc/s off resonance by rather more than 12 db., equivalent to two units of signal strength on the operating scale,

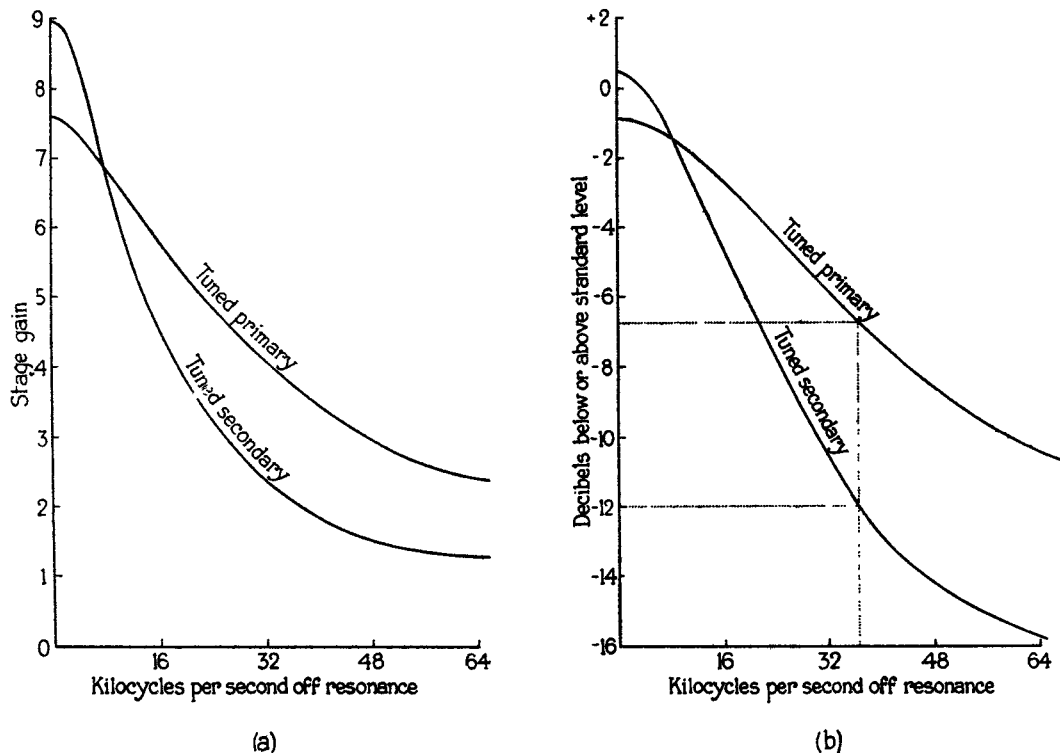


FIG. 31, CHAP. XI.—Comparison of selectivity of R.F. amplifying stages.

whereas the tuned primary circuit attenuates it by only 6 db. The relative selectivity of the two arrangements will be of the same order over the whole band covered by the tuning condenser, although the actual selectivity will be different, because the resistance of the tuned circuit increases with frequency.

INPUT ADMITTANCE

Components of input admittance

53. It has hitherto been assumed that if the grid and filament of a triode are connected to the two ends of an impedance, the effect is to add to the latter a small effective capacitance in parallel. In practice, the impedance added to the input circuit is of a more complex nature than this; in discussing the various forms it may take it is convenient to refer to the input admittance of the triode, i.e. the reciprocal of its apparent impedance, measured between grid and filament.

This input admittance will be denoted by $Y_i = \frac{1}{Z_i}$ and in general consists of a susceptance

$B_i = \omega C_i$ and a conductance $G_i = \frac{1}{R_i}$ in parallel, ω being 2π times the frequency of the applied

E.M.F. as usual. The input conductance G_i consists of three portions, first the leakage conductance G_l due to imperfect insulation between grid and filament. The insulation resistance will rarely exceed 50 megohms and may be much less, e.g. if a grid leak resistance is fitted. The lower limit of the conductance is therefore of the order of 2×10^{-8} siemens (mho). Second, a conductance due to the flow of an electron convection current between the filament and grid of the valve, commonly referred to as grid current. This conductance will be denoted by g_g and the corresponding resistance by r_g . Third, a conductance G_m due to the finite admittance Y_{ag} of the grid-anode path of the valve. The input susceptance B_i consists of two portions, first, that due to the grid-filament capacitance C_{gf} and second, that due to the finite admittance of the grid-anode path; this will be denoted by B_m .

The grid-anode admittance

54. (i) This consists of a conductance $G_{ag} = \frac{1}{R_{ag}}$ and a susceptance $B_{ag} = \omega C_{ag}$ in parallel,

and serves in effect to couple the anode and grid circuits of the valve. It is therefore responsible for a transference of energy from the anode to the grid circuit or vice versa, according to the operating conditions. The phenomena which are directly due to this coupling are collectively referred to as the Miller effect. In ordinary triodes the grid-anode capacitance usually lies between 2 and 10 $\mu\mu F$, while the resistance R_{ag} rarely exceeds 50 megohms. For theoretical purposes it is often convenient to consider this resistance to be infinite, but the effect of its finite value is of importance at audio frequencies. It will be seen later that the input conductance due to the Miller effect may in certain circumstances be of negative sign. In Chapter X it was shown that if the anode circuit of a simple receiver is coupled to the tuned input circuit by mutual inductance, an effective resistance is transferred to the input circuit, and further that this transferred resistance may be either of positive sign, tending to increase the damping, or of negative sign, tending to reduce the damping, according to the sign of the mutual inductance. Since the grid-anode capacitance acts as a coupling between the input circuit and the anode load impedance its effect is also to transfer an effective resistance to the input circuit. With a resistive or capacitive load, the transferred resistance is positive, but if the load offers inductive reactance, the transferred resistance may be negative.

(ii) For ease of reference the notation explained above is collected in the following table.

Z_i	= total input impedance	$= \frac{1}{Y_i}$.
Y_i	= total input admittance	$= \sqrt{G_i^2 + B_i^2}$.
B_i	= total input susceptance	$= B_{gf} + B_m$.
G_i	= total input conductance	$= G_l + G_m + g_g$.
B_{gf}	= susceptance of grid-filament path	$= \omega C_{gf}$.
G_l	= leakage conductance of grid-filament path.	
g_g	= internal grid-filament conductance of valve	$= \frac{1}{r_g}$.
G_m	= effective input conductance due to Miller effect.	
B_m	= effective input susceptance due to Miller effect.	
Y_{ag}	= admittance of grid-anode path	$= \sqrt{G_{ag}^2 + B_{ag}^2}$.
G_{ag}	= conductance of grid-anode path	$= \frac{1}{R_{ag}}$.
B_{ag}	= susceptance of grid-anode path	$= \omega C_{ag}$.

CHAPTER XI.—PARA. 55

In subsequent paragraphs, the following notation will also occur.

- G_d = dynamic conductance of input circuit alone.
- G'_d = total dynamic conductance of input circuit = $G_d + G_i$.
- χ = magnification of input circuit alone.
- χ' = effective magnification of input circuit.
- L_o = effective inductance of anode circuit load.
- R_{do} = dynamic resistance of tuned anode circuit.

Positive and negative reaction effects

55. (i) The effect of the input conductance of the valve upon the grid-filament circuit is most easily shown by numerical examples. Suppose the circuit to consist of an inductance of $160 \mu H$, a capacitance of $0.00025 \mu F$ and a resistance of 10 ohms, so that its dynamic resistance R_d is 64,000 ohms and its dynamic conductance G_d is 1.5625×10^{-5} siemens. If the valve has a positive input conductance $G_i = 10^{-5}$ siemens, the total conductance between grid and filament is

$$\begin{aligned} G'_d &= G_d + G_i \\ &= (1.5625 + 1) \times 10^{-5} \\ &= 2.5625 \times 10^{-5} \text{ siemens.} \end{aligned}$$

The effective dynamic resistance of the input circuit is therefore reduced from R_d to R'_d , where

$$\begin{aligned} R'_d &= \frac{10^5}{2.5625} \text{ ohms} \\ &= 39,000 \text{ ohms} \end{aligned}$$

and the effective magnification is correspondingly less. Since $R_d = \frac{\omega^2 L^2}{R}$ and the magnification

is $\chi = \frac{\omega L}{R}$, $\chi = \frac{R_d}{\omega L}$. The magnification of the tuned circuit alone is therefore $\frac{64,000}{800} = 80$,

but when the input conductance of the valve is connected in parallel, the magnification is reduced to

$$\begin{aligned} \chi' &= \frac{R'_d}{\omega L} \\ &= \frac{39,000}{80} \\ &= 48.8 \text{ (approx.).} \end{aligned}$$

For a given induced E.M.F. the effect of the input conductance in this particular instance, is to reduce the grid-filament P.D. in the ratio of 80 to 48.8.

(ii) Now suppose the input conductance to be of the same magnitude, namely 10^{-5} siemens, but of negative sign. The effective conductance between grid and filament then becomes

$$\begin{aligned} G'_d &= G_d + G_i \\ &= (1.5625 - 1) 10^{-5} \\ &= 5.625 \times 10^{-6} \text{ siemens,} \end{aligned}$$

and the effective dynamic resistance is increased to 177,600 ohms. The circuit magnification becomes $\chi' = 222$, so that for a given induced E.M.F., the grid-filament P.D. is nearly 2.8 times that obtained when the input conductance is zero. This increase of circuit magnification, giving rise to an increased signal strength for the same induced E.M.F., is thus responsible for the phenomenon known as regenerative amplification.

Effect of grid current

56. (i) In fig. 32a, e_g is the instantaneous E.M.F. of a source of alternating voltage of internal impedance Z , and T a triode valve, the grid of which is maintained at such a mean negative potential E_b with respect to the filament, that no grid current flows during any portion of the cycle. These operating conditions are shown in fig. 32b. In these circumstances no current will be supplied by the source, and the alternating voltage V_g between grid and filament will be equal to the E.M.F. of the source, i.e. $v_g = e_g$; the triode may therefore be said to impose no load upon the generator.

(ii) Now suppose the bias voltage to be removed and grid current to flow during the whole of the cycle, the operating conditions being those of fig. 32c. The electron stream between filament and grid will now be subjected to an alternating voltage and a pulsating movement will be superimposed upon the mean steady progression of the electrons. Work must be done by

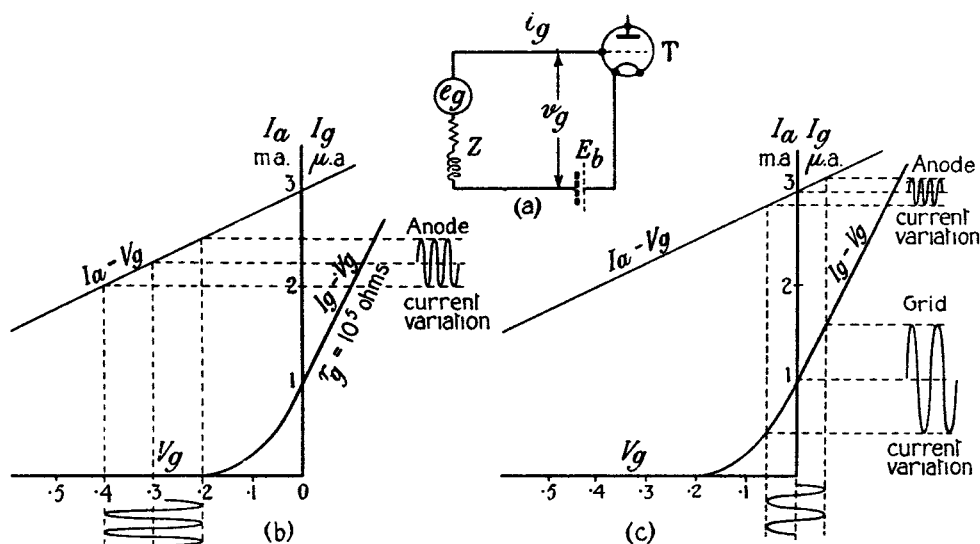


FIG. 32, CHAP. XI.—Effect of grid current.

the source in order to cause this pulsation, and its power output will be equal to the rate at which work is performed. The pulsation of the electrons within the valve must result in a similar motion of the electrons in the external circuit, so that an alternating current i_g (I_g , R.M.S.) may be considered to exist. There is therefore an internal voltage drop, $i_g Z$, in the generator, and the grid-filament P.D. will be less than the E.M.F., i.e. $v_g = e_g - i_g Z$. Electrical energy is converted into heat owing to the motion of the electrons, and the power expended in the valve

is equal to $I_g^2 r_g$ watts, where $r_g = \frac{1}{g_g}$ and g_g is the slope of the $I_g - V_g$ curve. In fig. 32 r_g is approximately equal to 10^5 ohms, hence if the other components of the input conductance are zero, the effect of this grid current upon a tuned input circuit in which $L = 160 \mu H$, $C = 0.00025 \mu F$, $R = 10$ ohms, is precisely as calculated in paragraph 55.

(iii) If the operating conditions are intermediate between those of fig. 32b and fig. 32c so that grid current flows during only a portion of the cycle, the input conductance will not be constant, and the wave-form of the grid-filament P.D. will not be identical with that of the E.M.F.

CHAPTER XI.—PARA. 57

of the source. This must occur to some extent in all cases where grid current is allowed to flow, since the $I_g - V_g$ characteristic always possesses considerable curvature. In general then, it may be said that in addition to the reduction in grid-filament voltage, amplitude distortion must arise if grid current is permitted.

Effect of grid-anode admittance

57. (i) In dealing with the portion $Y_m = \sqrt{G_m^2 + B_m^2}$, of the input admittance which is due to the finite impedance of the grid-anode path, it is convenient to assume that the operating conditions are those of fig. 32a, i.e. that no grid current flows. The circuit diagram of the amplifier is then as given in fig. 33a in which the grid-filament voltage is $v_g = \mathcal{V}_g \sin \omega t$, its

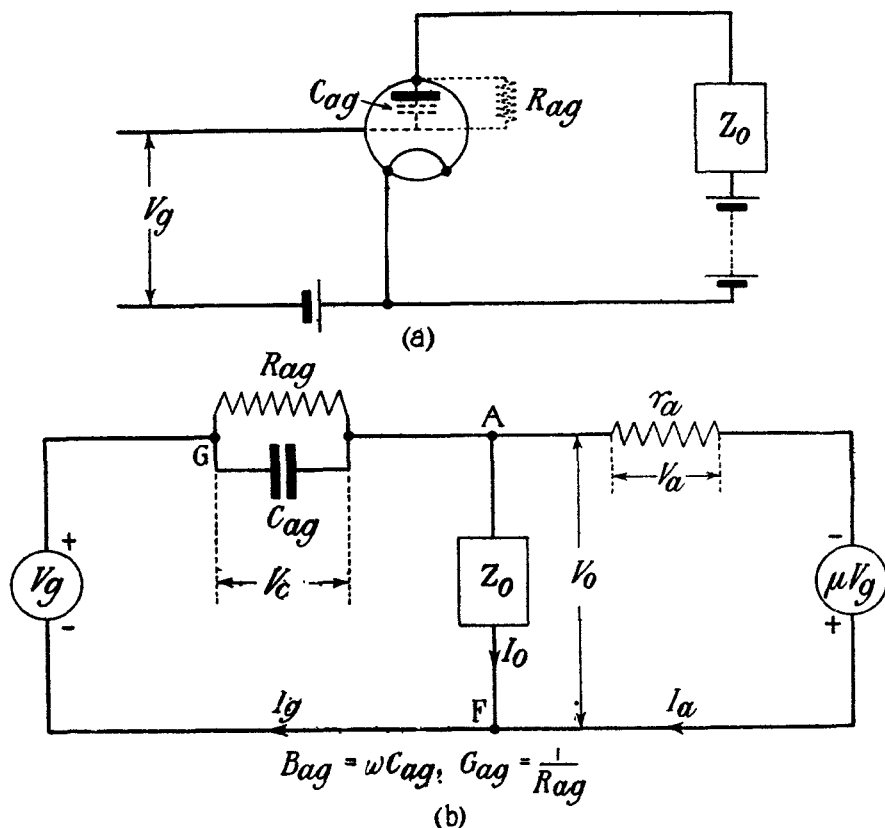


FIG. 33, CHAP. XI.—Anode-grid admittance of amplifier and equivalent circuit.

R.M.S. value being V_g . The load impedance is denoted by Z_0 , the grid-anode capacitance by C_{ag} , and the grid-anode conductance by $G_{ag} = \frac{1}{R_{ag}}$. The electrical equivalent of the circuit is

given in fig. 33b in which the input voltage is V_g and an amplified voltage μV_g is assumed to act in series with the anode A.C. resistance r_a . It is now obvious that owing to the admittance Y_{ag} the input voltage is able to supply current directly to the load impedance, while the equivalent generator μV_g is able to supply current to the input circuit.

(ii) In certain numerical examples, the triode will be assumed to possess the following constants, viz. :—

$$\begin{aligned}\mu &= 10 \\ r_a &= 2 \times 10^4 \text{ ohms} \\ C_{ag} &= 5 \times 10^{-12} \text{ Farad.} \\ R_{ag} &= 5 \times 10^7 \text{ ohms } (G_{ag} = 2 \times 10^{-6} \text{ siemens}).\end{aligned}$$

This will be referred to as the typical triode.

(iii) When B_{ag} is very much larger than G_{ag} , i.e. at frequencies above about 8,000 cycles per second, G_{ag} may be ignored; at frequencies in the region of 600 cycles per second, B_{ag} is approximately equal to G_{ag} and both must be taken into account. At frequencies below about 400 cycles per second, the grid-anode admittance may be approximately represented by its conductance only. The current flowing through the generator V_g and admittance Y_{ag} will be referred to as the grid current, but must not be confused with the true grid-filament current previously discussed. Its instantaneous value is i_g and its R.M.S. value I_g . The corresponding anode current is i_a and the current through the load impedance i_o , their respective R.M.S. values being I_a and I_o . The conventional signs (+ and -) on the respective generators are inserted merely to show their relative phase.

58. From fig. 33 certain deductions may be made by inspection. If G_{ag} is zero, and the load impedance is a loss-free acceptor circuit for the applied frequency, $I_a = \frac{\mu V_g}{r_a}$ and $I_g = \omega C_{ag} V_g$, i.e. the input admittance Y_m is equal to B_{ag} . If the anode load is of this nature, then, the input admittance is purely capacitive and imposes no damping upon the input circuit. On the other hand, if the load impedance is a loss-free rejector circuit, the load current I_o is

$$\text{zero and } I_g = I_a = \frac{(\mu + 1) V_g}{\sqrt{r_a^2 + \left(\frac{1}{\omega C_{ag}}\right)^2}}$$

Under these conditions

$$Y_m = \frac{\mu + 1}{\sqrt{r_a^2 + \left(\frac{1}{\omega C_{ag}}\right)^2}} = \frac{(\mu + 1) B_{ag}}{\sqrt{g_a^2 + B_{ag}^2}} g_a.$$

Y_m is easily resolved into its conductive and susceptive components,

$$G_m = \frac{(\mu + 1) r_a}{r_a^2 + \left(\frac{1}{\omega C_{ag}}\right)^2} = \frac{(\mu + 1) B_{ag}^2}{g_a^2 + B_{ag}^2} g_a$$

and

$$B_m = \frac{(\mu + 1) \frac{1}{\omega C_{ag}}}{r_a^2 + \left(\frac{1}{\omega C_{ag}}\right)^2} = \frac{(\mu + 1) g_a^2}{g_a^2 + B_{ag}^2} B_{ag}.$$

CHAPTER XI.—PARAS. 59-60

Example.—Under the above conditions, find the input conductance and susceptance of the typical triode at 796 kc/s ($\omega = 5 \times 10^6$).

$$\begin{aligned} B_{ag} &= \omega C_{ag} = 5 \times 10^6 \times 5 \times 10^{-12} \\ &= 25 \times 10^{-6} \\ \frac{1}{\omega C_{ag}} &= 4 \times 10^4 \\ G_m &= \frac{(\mu + 1) r_a}{r_a^2 + \left(\frac{1}{\omega C_{ag}}\right)^2} \\ &= \frac{11 \times 2 \times 10^4}{(2 \times 10^4)^2 + (4 \times 10^4)^2} \\ &= 1.1 \times 10^{-4} \text{ siemens} \\ B_m &= \frac{(\mu + 1) \frac{1}{\omega C_{ag}}}{r_a^2 + \left(\frac{1}{\omega C_{ag}}\right)^2} \\ &= \frac{11 \times 4 \times 10^4}{20 \times 10^8} \\ &= 2.2 \times 10^{-4} \text{ siemens.} \end{aligned}$$

Finite, purely resistive load

59. If, as before, we assume the grid-anode conductance to be zero, and the anode load is a pure resistance, r_o , the values of G_m and B_m are given by the following expressions.

$$\begin{aligned} G_m &= \frac{A + 1}{Z^2} \quad R = \frac{(A + 1) B_{ag}^2}{G^2 + B_{ag}^2} G \\ B_m &= \frac{A + 1}{Z^2} \frac{1}{B_{ag}} = \frac{(A + 1) G^2}{G^2 + B_{ag}^2} B_{ag} \\ \text{where } A &= \frac{r_o}{r_a + r_o} \mu \\ R &= \frac{r_o r_a}{r_a + r_o} = \frac{1}{G} \\ Z^2 &= R^2 + \left(\frac{1}{\omega C_{ag}}\right)^2 \end{aligned}$$

It will be observed that the above equations are of the same form as for an infinite load, but are modified by the substitution of the V.A.F. (A) for the amplification factor (μ), and the conductance (G) of the valve and load in parallel for that of the valve alone.

60. The effects of a resistive load are shown by a vector diagram in fig. 34. The loads on the two sources of supply are shown separately in fig. 34a. Looking into the circuit from the generator V_g the load impedance is that of the capacitance C_{ag} in series with the parallel combination of r_a and r_o , i.e. R . The current through C_{ag} due to this voltage is

$$I_1 = \frac{V_g}{\sqrt{R^2 + \left(\frac{1}{\omega C_{ag}}\right)^2}}$$

In the vector diagram (fig. 34b), I_1 is leading on V_g by nearly 90° , i.e. B_{ag} is assumed to be very much smaller than G . From the point of view of the generator μV_g , the load is composed of the resistance r_a in series with the parallel combination of r_o and C_{ag} . The current due to the voltage μV_g is denoted by I_2 , its components being I_c flowing through C_{ag} and I_o flowing through the load resistance r_o ; I_2 will lead on μV_g by an angle much less than 90° , because I_c will lead on μV_g by 90° and I_o will be in phase with μV_g . We are chiefly concerned with the component I_c , which combines with I_1 to give the total current I_g through the capacitance C_{ag} . Since I_g is the vector sum of I_c and I_1 , it leads on V_g by an angle θ which is less than 90° . It must have a component $I_g \cos \theta$ in phase with V_g and the power dissipated in the input circuit will be $I_g V_g \cos \theta$. This power is always small because in all practical amplifiers B_{ag} is small compared with G so that the angle θ is very nearly 90° . The effect of the grid-anode susceptance is therefore merely to increase the input capacitance by an amount which is less than $(A + 1) C_{ag}$. At very low audio frequencies, the admittance G_{ag} of the grid-anode path may be of considerable

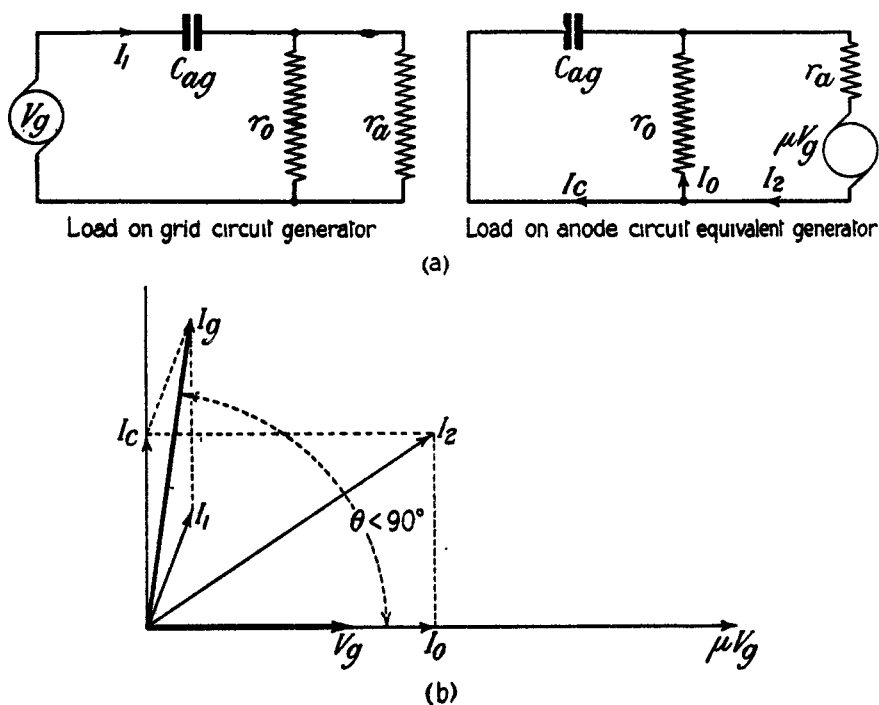


FIG. 34, CHAP. XI.—Effect of finite, resistive load.

importance. When B_{ag} is very small compared with G_{ag} the input susceptance B_m is negligible, while the input conductance G_m approaches the value $(A + 1) G_{ag}$. Thus at 157 cycles per second, the grid-anode susceptance of the typical triode is $B_{ag} = 2\pi \times 157 \times 5 \times 10^{-12} = 5 \times 10^{-9}$

and its conductance G_{ag} is 2×10^{-8} , i.e. $G_{ag} = 4 B_{ag}$. Let $r_o = r_a$, so that $A = \frac{\mu}{2} = 5$,

$A + 1 = 6$. Then $G_m = 6 \times 2 \times 10^{-8} = 1.2 \times 10^{-7}$ siemens. Thus the input conductance is equivalent in effect to a leak of about 8 megohms. This may not appear to be serious, but the effect may be appreciable if G_{ag} is allowed to assume a high value owing to an accumulation of dirt or moisture between grid and anode connections.

Capacitive load

61. If the load impedance offers capacitive reactance, e.g. if it consists of a resistance and a condenser in parallel, the input admittance becomes that of a positive conductance and a capacitive susceptance in parallel. The vector diagram for a purely capacitive load is given in fig. 35. Taking μV_g , the total voltage acting in the anode circuit, as the datum vector, the anode current I_a leads on μV_g by an angle less than 90° . The anode-filament P.D. V_a will be equal in magnitude to $I_a r_a$ and will be in phase with I_a . The P.D. V_o across the load will be equal to $\mu V_g - V_a$, vectorially, and the load current I_o will lead on the voltage V_o by 90° . The grid current I_g is equal to $I_o - I_a$ vectorially, and is therefore less than 90° ahead of V_o , i.e. very much less than 90° ahead of V_g . As a result, the input admittance has a large

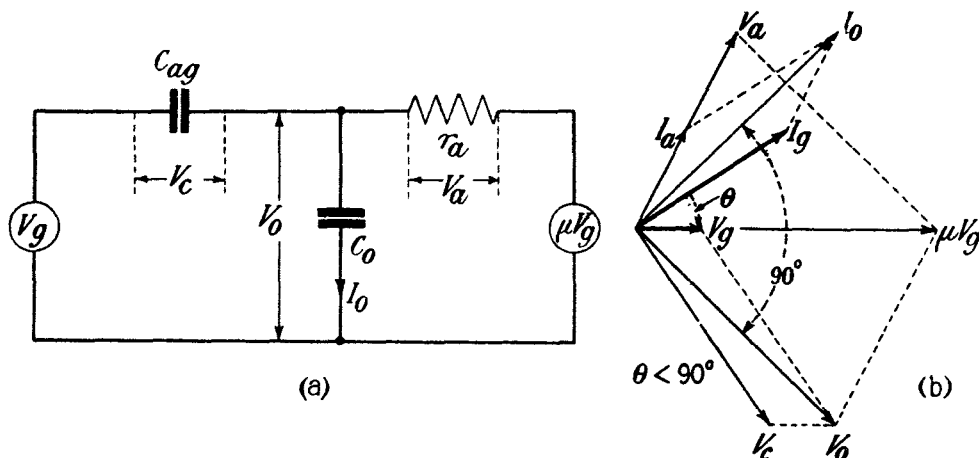


FIG. 35, CHAP. XI.—Effect of finite, capacitive load.

conductance component, and the valve may, in certain circumstances, impose a very heavy load upon the input circuit. The maximum value of the input conductance (assuming G_{ag} to be zero) is

$$G_m (\text{max.}) = \frac{\mu}{2} B_{ag}$$

and is obtained with a load reactance equal in magnitude to r_a .

Inductive load

62. When the anode load offers inductive reactance, the input conductance may become negative in sign, and the manner in which this comes about may again be explained by means of a vector diagram. Referring to fig. 36 and taking the vector μV_g as datum, the anode current I_a will lag on this by some angle less than 90° , and the anode-filament P.D., $V_a = I_a r_a$, will be in phase with I_a . The P.D. V_o across the load will be equal to $\mu V_g - V_a$, vectorially, while the load current I_o lags on V_o by nearly 90° . The grid current I_g is equal to $I_a - I_o$, vectorially, while the grid-anode P.D. V_c is the vector sum of V_g and V_o . Provided that G_{ag} is negligible, therefore, I_g leads on the voltage V_c by 90° . Since V_c must lead on V_g , I_g must lead on V_g by an angle θ greater than 90° . As the power input to the grid circuit is $V_g I_g \cos \theta$, and $\cos \theta$ is negative if θ lies between 90° and 270° , it follows that under these conditions the "generator" V_g is receiving power from the circuit instead of supplying power. The maximum negative value of the input conductance (assuming G_{ag} to be zero) is

$$G_m (\text{max.}) = -\frac{\mu}{2} B_{ag}$$

and is obtained with a load reactance equal in magnitude to r_a .

63. Whereas with a capacitive load the input conductance is invariably positive, with an inductive load it may be positive or negative. There is in fact, a kind of resonance effect between the effective load inductance L_o and the grid-anode capacitance C_{ag} ; at frequencies for which G_{ag} is negligible compared with B_{ag} ,

$$G_m = \frac{\frac{\omega^2 L_o^2}{r_a} (\mu + 1) - \frac{L_o \mu}{C_{ag} r_a}}{\left(\frac{L_o}{C_{ag} r_a}\right)^2 + \left(\omega L_o - \frac{1}{\omega C_{ag}}\right)^2} \quad (1)$$

The resonance effect occurs when

$$\omega^2 = \frac{1}{L_o C_{ag} \left(1 + \frac{1}{\mu}\right)} \quad (2)$$

and at this frequency $G_m = 0$. At lower frequencies G_m is negative and at higher frequencies is positive. The magnitude of the input conductance at audio frequencies may be illustrated numerically as follows.

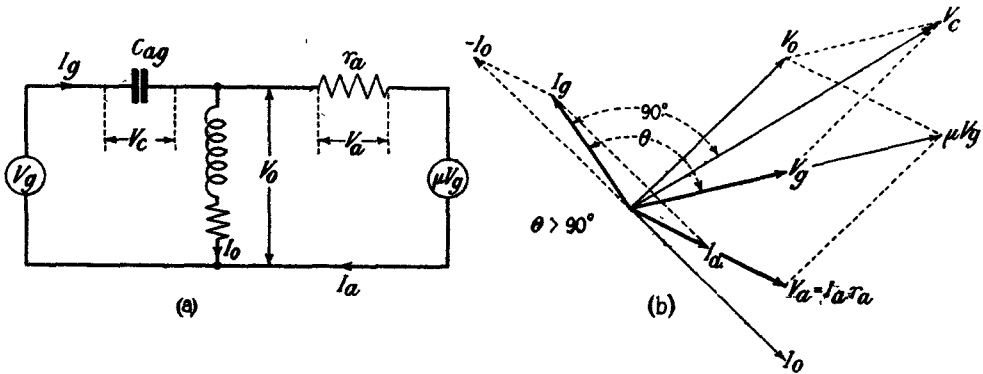


FIG. 36, CHAP. XI.—Effect of finite inductive load.

Example.—The anode load of a typical triode is an inductance of 50 henries. Assuming G_{ag} to be negligible, find the input conductance when $\omega = 5 \times 10^3$, $\omega = 5 \times 10^4$ and $\omega = 7 \times 10^4$.

(i)

$$\begin{aligned} \omega &= 5 \times 10^3 \\ \omega L_o &= 5 \times 10^3 \times 50 = 2.5 \times 10^5 \\ \frac{(\omega L_o)^2}{r_a} (\mu + 1) &= 6.25 \times 10^{10} \times 11 \times 5 \times 10^{-5} \\ &= 3.44 \times 10^7 \\ \frac{L_o}{C_{ag} r_a} &= 50 \times 2 \times 10^{11} \times 5 \times 10^{-5} \\ &= 5 \times 10^8 \\ \frac{\mu L_o}{C_{ag} r_a} &= 5 \times 10^9 \\ \frac{1}{\omega C_{ag}} &= 2 \times 10^{-4} \times 2 \times 10^{11} \\ &= 4 \times 10^7 \\ \omega L_o - \frac{1}{\omega C_{ag}} &= 2.5 \times 10^5 - 4 \times 10^7 \\ &\doteq -4 \times 10^7. \end{aligned}$$

CHAPTER XI.—PARA. 64

It is apparent that the positive term in the numerator is negligible compared with the negative term, and that the denominator is practically equal to $\left(\frac{L_o}{C_{ag} r_a}\right)^2$. Hence

$$\begin{aligned} G_m &= - \frac{\frac{\mu L_o}{C_{ag} r_a}}{\left(\frac{L_o}{C_{ag} r_a}\right)^2} \\ &= - \mu \frac{C_{ag} r_a}{L_o} \\ &= - 10 \times \frac{1}{5 \times 10^8} \\ &= - 2 \times 10^{-8} \text{ siemens.} \end{aligned}$$

(ii)

$$\begin{aligned} \omega &= 5 \times 10^4 \\ \omega L_o &= 2.5 \times 10^6 \\ \frac{(\omega L_o)^2}{r_a} (\mu + 1) &= 3.44 \times 10^9 \\ \frac{L_o}{C_{ag} r_a} &= 5 \times 10^8, \quad \frac{\mu L_o}{C_{ag} r_a} = 5 \times 10^9 \text{ as before.} \\ \frac{1}{\omega C_{ag}} &= 4 \times 10^6 \end{aligned}$$

$\left(\omega L_o - \frac{1}{\omega C_{ag}}\right)^2$ is again negligibly small compared with $\left(\frac{L_o}{C_{ag} r_a}\right)^2$

$$\begin{aligned} G_m &= \frac{(3.44 - 5) 10^9}{25 \times 10^6} \\ &= - .624 \times 10^{-8} \text{ siemens.} \end{aligned}$$

(iii)

$$\begin{aligned} \omega &= 7 \times 10^4 \\ \omega L_o &= 3.5 \times 10^6 \\ \frac{(\omega L_o)^2}{r_a} (\mu + 1) &= 6.7 \times 10^9 \end{aligned}$$

$\frac{L_o}{C_{ag} r_a}$ and $\frac{\mu L_o}{C_{ag} r_a}$ remain as before, and the denominator is to all intents and purposes equal to 25×10^{16}

$$\begin{aligned} G_m &= \frac{(6.7 - 5) 10^9}{25 \times 10^{16}} \\ &= + .684 \times 10^{-8} \text{ siemens.} \end{aligned}$$

64. At audio frequencies, no matter what the nature of the load may be, the input admittance consists of a capacitive susceptance, approximately equal to $(A + 1) B_{ag}$, and a conductance, nearly always positive. Its magnitude is governed largely by the conductance G_{ag} of the grid-anode path, which has hitherto generally been assumed to be negligible. For practical purposes, it may be said that the input conductance G_m is positive if G_{ag} exceeds $\frac{B_{ag}}{2}$,

which is nearly always true at audio frequencies. Under these conditions the magnitude of G_m is approximately equal to $\left(1 + \frac{A^2}{\mu}\right) G_{ag}$. At radio frequencies, the effect of the anode-grid conductance is negligible, and the input conductance is very nearly that given by equation 1, paragraph 63. Its order of magnitude will now be illustrated with reference to the typical triode.

Example.—The anode load of the typical triode is an inductance of $200 \mu H$. Find the input conductance when $\omega = 5 \times 10^6$.

$$\omega L_o = 5 \times 10^6 \times 200 \times 10^{-6} = 10^3$$

$$\frac{(\omega L_o)^2}{r_a} (\mu + 1) = 10^6 \times 11 \times 5 \times 10^{-5} \\ = 5.5 \times 10^3$$

$$\frac{L_o}{C_{ag} r_a} = 200 \times 10^{-6} \times 2 \times 10^{11} \times 5 \times 10^{-5} \\ = 2 \times 10^3$$

$$\frac{\mu L_o}{C_{ag} r_a} = 2 \times 10^4$$

Obviously $\frac{\mu L_o}{C_{ag} r_a}$ is very much larger than $\frac{(\omega L_o)^2}{r_a} (\mu + 1)$.

$$\frac{1}{\omega C_{ag}} = \frac{1}{5 \times 10^6 \times 5 \times 10^{-12}} = 4 \times 10^4$$

$$\left(\omega L_o - \frac{1}{\omega C_{ag}}\right)^2 \div \left(\frac{1}{\omega C_{ag}}\right)^2 = 8 \times 10^8$$

and is large compared with $\frac{L_o}{C_{ag} r_a}$.

Hence the input conductance is approximately

$$G_m = - \frac{2 \times 10^4}{8 \times 10^8} \\ = - 2.5 \times 10^{-5} \text{ siemens.}$$

The corresponding input shunt resistance is $- 4 \times 10^4$ ohms.

Self-oscillation due to negative input conductance

65. The effect of this negative input conductance upon the input circuit depends upon the dynamic conductance of the circuit itself. If the latter is that considered in paragraph 55, having a dynamic conductance of 1.5625×10^{-5} siemens, the total conductance will be

$$G'_d = G_d + G_i = (1.5625 - 2.5) \times 10^{-5} \\ = - .9375 \times 10^{-5} \text{ siemens.}$$

The total conductance between grid and filament is therefore negative. Although this condition may exist momentarily, it cannot persist. Suppose it to be true at the instant of switching on the H.T. supply to the valve, bearing in mind that the mere acceleration of electrons consequent upon this action must give rise to an oscillation at some frequency or other. The valve will then supply power to the input circuit and, since the effective resistance of the latter is momentarily negative, it will set up a large circulating current and a corresponding grid-filament P.D. This large grid swing will give rise to grid current and the input conductance will become less negative, while since the D.C. resistance of the input circuit cannot be zero, some mean negative grid bias will be developed. The operating point of the dynamic $I_a - V_a$ curve will therefore

CHAPTER XI.—PARA. 66

move to a point of greater curvature, i.e. where the anode A.C. resistance is greater. Thus the valve itself adjusts the various operating parameters in such a manner that, although an oscillatory current is maintained in the input circuit, the total grid-filament conductance becomes zero, while the power dissipated in the valve and its associated circuits is exactly equal to that supplied by the H.T. source; the grid voltage swing is then maintained at the magnitude which it possessed at the instant of stabilization. This, briefly, is the mechanism by which an oscillation is maintained as a result of the capacitive coupling between grid and anode.

Input conductance with tuned anode

66. Even with an inductive load, however, the input conductance is not necessarily negative. The input conductance is positive if

$$\omega^2 > \frac{1}{L_o C_{ag} \left(1 + \frac{1}{\mu}\right)}$$

as already explained. It is of interest to find the magnitude of the load inductance which will make the input conductance positive; e.g. at 796 kc/s the input conductance is positive if

$$\begin{aligned} L_o &> \frac{1}{\omega^2 C_{ag} \left(1 + \frac{1}{\mu}\right)} \\ &> \frac{10^{12}}{25 \times 10^{12} \times 5 \times 1.1}, \end{aligned}$$

i.e. if

$$L_o > 7,260 \mu H.$$

All practical radio-frequency amplifiers may be reduced to an equivalent tuned anode circuit by the methods of paras. 49 *et seq.* and at the resonant frequency of the tuned circuit the load is purely resistive. At frequencies very close to resonance, however, the load becomes highly reactive—inductive if the circuit is tuned to a frequency higher than that applied and vice versa. Thus, if the anode circuit has an inductance L of $200 \mu H$ in parallel with a capacitance C of $.0002 \mu F$, and negligible resistance, its reactance at its resonant frequency is infinite; at a

slightly lower frequency, e.g. $\omega = 4.9 \times 10^6$, its reactance is $X_o = \frac{\omega L}{1 - \omega^2 LC}$ ohms, and is inductive if the sign of this quantity is positive. The reactance at $\omega = 4.9 \times 10^6$ is

$$\begin{aligned} X_o &= \frac{4.9 \times 200 \times 10^6 \times 10^{-6}}{1 - 4.9^2 \times 10^{12} \times 200 \times 10^{-6} \times .0002 \times 10^{-6}} \\ &= \frac{980}{1 - .96} \\ &= 24,500 \text{ ohms,} \end{aligned}$$

and is positive. It may therefore be represented by an equivalent inductance L_o , where

$$\begin{aligned} L_o &= \frac{24,500}{\omega} \\ &= 4,900 \mu H \end{aligned}$$

With the typical triode, therefore, the input conductance at this frequency would be approximately $-\frac{\mu C_{ag} r_a}{L_o}$ or $-\frac{1}{200}$ siemens and if the input circuit is tuned to the frequency corresponding to $\omega = 4.9 \times 10^6$ the mere switching on of the power supply is sufficient to cause self-oscillation as explained in paragraph 65.

67. On the other hand, if the applied frequency is slightly higher than the resonant frequency of the anode circuit, e.g. if $\omega = 5.1 \times 10^6$ the load reactance is approximately of the same magnitude as in the case just discussed, but of positive sign. The effective dynamic resistance of the input circuit is then less than 200 ohms. The following example is illustrative of the conditions which may exist at very high radio frequencies.

Example.—(i) The typical triode is required to amplify a signal at the frequency corresponding to $\omega = 4.9 \times 10^7$, the equivalent tuned anode having the constants $L = 20 \mu H$, $C = 20 \mu \mu F$, $R = 10$ ohms. Find the input conductance.

The resonant frequency of the tuned anode is $\frac{5 \times 10^7}{2\pi}$, and its reactance at the applied frequency is

$$\begin{aligned} X_o &= \frac{\omega L}{1 - \omega^2 LC} \\ &= 24,500 \text{ ohms} \\ L_o &= 490 \mu H \\ \frac{(\omega L_o)^2}{r_a} (\mu + 1) &= (2.45 \times 10^4)^2 \times 11 \times 5 \times 10^{-5} \\ &= 330,000 \\ \frac{L_o}{C_{ag} r_a} &= 4.9 \times 10^{-4} \times 2 \times 10^{11} \times 5 \times 10^{-5} \\ &= 4,900 \\ \mu \frac{L_o}{C_{ag} r_a} &= 49,000. \\ \frac{1}{\omega C_{ag}} &= \frac{2 \times 10^{11}}{4.9 \times 10^7} \\ &= 4,070 \\ \omega L_o - \frac{1}{\omega C_{ag}} &\doteq 2 \times 10^4 \\ G_m &= \frac{(330 - 49) 10^3}{(\cdot 49 \times 10^4)^2 + (2 \times 10^4)^2} \\ &= \frac{281}{4.24 \times 10^5} \\ &= 6 \times 10^{-4} \text{ siemens,} \end{aligned}$$

which is equivalent to a leak of 1,667 ohms across the input circuit.

(ii) If the aerial circuit has an inductance L_A of $40 \mu H$ and a resistance R_A of 100 ohms find the overall gain of the aerial and amplifying stage.

The dynamic resistance of the aerial circuit is $\frac{\omega^2 L_A^2}{R_A} \doteq 40,000$ ohms, and its magnification, $x \doteq 20$. The total dynamic conductance of the input circuit is

$$\begin{aligned} G'_d &= \frac{1}{4 \times 10^4} + \frac{6}{10^4} \\ &= \frac{6.25}{10^4} \\ R'_d &= \frac{10^4}{6.25} = 1,600 \text{ ohms.} \end{aligned}$$

CHAPTER XI.—PARAS. 68-69

The effective magnification of the input circuit is therefore

$$\begin{aligned} z &= \frac{R'_d}{\omega L} = \frac{1,600}{5 \times 10^7 \times 40 \times 10^{-6}} \\ &= .8. \end{aligned}$$

The V.A.F. of the amplifier alone is

$$\frac{R_{d0}}{R_{d0} + r_a} \mu$$

where R_{d0} is the dynamic resistance of the anode circuit and is equal to 100,000 ohms.

$$\begin{aligned} \text{V.A.F.} &= \frac{100,000 \times 10}{120,000} \\ &= 8.7 \end{aligned}$$

and the overall gain only $.8 \times 8.7 \div 7$ which is of course much less than the magnification of the aerial circuit.

68. The introduction of an amplifier stage may therefore actually reduce the voltage available at the detector. It is often thought that the valve fails to act as an amplifier at high frequencies, but the above example shows that this is not the case. The valve performs its amplification in the manner predicted by the theory discussed in paras. 45 *et seq.*, but fails to give an overall gain owing to the damping it imposes upon its input circuit. If however the anode circuit is very slightly detuned so that its resonant frequency is very slightly higher than the signal frequency, its input conductance becomes sufficiently negative to maintain the input circuit in oscillation. In practice it is impossible so to adjust the anode circuit that stable amplification is obtained, unless the input and output circuits are of very low dynamic resistance. In these conditions, although the amplifier gain may be appreciable, the overall gain is negligible. An amplifier in which stability is attained by the employment of a high ratio of capacitance to inductance, giving low dynamic resistance and negligible overall gain, is often referred to as a buffer amplifier.

Use of screen-grid valves

69. (i) Unless special precautions are adopted, then, it is almost impossible to maintain a tuned radio-frequency triode amplifier stage in a non-oscillatory condition. Although at first sight it would appear possible to adjust both input and output circuits to the same frequency, so that they offer dynamic resistance only, and thus to ensure a positive input conductance, this is not so in practice. One complication which arises is that the two circuits are capacitance-coupled as in fig. 30, Chapter VI, and therefore possess two resonant frequencies. If the circuit constants are (L_1, C_1) , (L_2, C_2) , and $L_1 \times C_1 = L_2 \times C_2 = LC$ we have

$$f_1 = \frac{1}{2\pi\sqrt{LC}}$$

and to this frequency both circuits are practically non-reactive. The other resonant frequency is

$$\begin{aligned} f_2 &= \frac{1}{2\pi\sqrt{LC\left(1 + \frac{C_{ag}}{C_1} + \frac{C_{ag}}{C_2}\right)}} \\ &= \frac{f_1}{\sqrt{\left(1 + \frac{C_{ag}}{C_1} + \frac{C_{ag}}{C_2}\right)}} \end{aligned}$$

which is obviously less than f_1 , hence both input and output circuits offer inductive reactance at this frequency. The input conductance will therefore be negative, and if the positive grid-filament conductance is sufficiently low the valve will maintain oscillations at this frequency.

(ii) Owing partly to its low amplification factor, but chiefly to the difficulty of preventing self-oscillation, the triode valve is rarely used for radio-frequency amplification in modern receivers. Either the screen-grid valve, or one of its later developments, e.g. the variable- μ screen-grid valve and the radio-frequency pentode, is invariably adopted. The variable- μ valve is chiefly used where it is necessary to adopt some form of automatic radio-frequency gain control, and is again referred to in Chapter XII. The object of the radio-frequency pentode is to avoid the effects of secondary emission and so enable a larger anode voltage swing than is permissible with screen-grid valves. It is therefore capable of giving a greater output voltage than its prototype. The variable- μ pentode combines both the above features and is largely used in broadcast receivers. The gain and stability of the screen-grid valve will now be dealt with, the discussion being equally applicable to valves of later development.

Gain and stability

70. The principal features and characteristic curves of the screen-grid valve were discussed in chapter VIII. It was there stated that the introduction of the screening electrode reduces the effective grid-anode capacitance to only a fraction of a micro-microfarad and the input admittance is therefore very much less than that of a triode. Under normal operating conditions, namely with a screen voltage of about two-thirds the anode voltage, and with the grid swing limited to rather less than two volts, the equivalent circuit (and therefore the method of computing stage gain) is identical with that of a triode amplifier. The S.G. valve is not inherently stable in operation as is sometimes thought, for although its anode-grid capacitance is low it is not zero, and there is an upper limit to the magnitude of the dynamic load resistance R_{do} on this account. The input conductance becomes negative at a frequency very near to resonance, its minimum value being approximately

$$G_m = - \frac{\omega C_{ag} A}{2} = - \frac{\omega C_{ag}}{2} \times \frac{g_m}{\frac{1}{R_{do}} + \frac{1}{r_a}}$$

Instability ensues if the total grid-filament conductance is zero, so that if G_d is the admittance of the input circuit in the absence of the valve the limiting condition for perfect stability is clearly

$$G_d + G_m = 0.$$

Inserting the minimum value of G_m

$$G_d - \frac{\omega C_{ag} g_m}{2 \left(\frac{1}{R_{do}} + \frac{1}{r_a} \right)} = 0$$

$$G_d \left(\frac{1}{R_{do}} + \frac{1}{r_a} \right) = \frac{\omega C_{ag} g_m}{2}$$

If for simplicity it is assumed that the input and output circuits are of similar design, so that

$$G_{do} = \frac{1}{R_{do}} = \frac{1}{R_d},$$

$$\frac{1}{R_d} \left(\frac{1}{R_d} + \frac{1}{r_a} \right) = \frac{\omega C_{ag} g_m}{2}.$$

This is the fundamental formula for use in calculations on stability. As in practice r_a is always very much larger than R_d , it is permissible to use a further approximation; provided

$$\frac{1}{R_d} \gg \frac{1}{r_a}, \quad \frac{1}{R_d} \left(\frac{1}{R_d} + \frac{1}{r_a} \right) \doteq \frac{1}{R_d^2}, \text{ and}$$

$$\frac{1}{R_d^2} = \frac{\omega C_{ag} g_m}{2}$$

CHAPTER XI.—PARAS. 71-72

The maximum permissible dynamic resistance with which the circuit will remain stable is therefore

$$R_d (\text{max.}) = \sqrt{\frac{2}{\omega C_{ag} g_m}}.$$

71. As an example of the use of this equation, we may calculate the maximum permissible dynamic resistance at 796 kc/s ($\omega = 5 \times 10^6$) for a typical battery-operated S.G. valve ($g_m = 1$ milliamperes per volt, $r_a = 300,000$ ohms, $C_{ag} = .005 \mu\mu F$.)

$$\begin{aligned} R_d (\text{max.}) &= \sqrt{\frac{2}{5 \times 10^6 \times 5 \times 10^{-15} \times 10^3}} \\ &= 283,000 \text{ ohms.} \end{aligned}$$

As the frequency is increased the permissible value of R_d becomes smaller, e.g. at $\omega = 5 \times 10^7$ it is only 90,000 ohms. Such high dynamic resistances are not ordinarily obtainable in the absence of regenerative amplification from the succeeding stage, and neglecting this complication the above assumption, $R_d < r_a$ invariably applies. The V.A.F. may then be reduced to a very simple expression, i.e.

$$\text{V.A.F.} = \frac{\mu R_d}{r_a} = R_d g_m$$

and with the maximum permissible value of R_d

$$\text{V.A.F.} = g_m \sqrt{\frac{2}{\omega C_{ag} g_m}} = \sqrt{\frac{2 g_m}{\omega C_{ag}}}.$$

At $\omega = 5 \times 10^6$ the maximum V.A.F. is about 200 and at $\omega = 5 \times 10^7$ is about 70. In practice the gain seldom approaches such high values without a tendency to instability, owing to the effects of other forms of coupling between input and output circuits. In any event it is absolutely essential that the magnetic and electrostatic screening shall be of a very high order if gains comparable with the theoretical maxima are sought. The selectivity of the S.G. amplifier is theoretically higher than that of a triode with an identical anode circuit, because the damping due to the anode A.C. resistance of the valve is much less. Compared with the tuned anode circuit, tuned transformer coupling does not improve the selectivity to the same extent as in triode amplification for the same reason. In R/T reception the improvement in selectivity is only achieved when the input circuit has a high degree of initial selectivity, owing to an effect known as cross-modulation which will receive attention in Chapter XII.

Multi-stage amplifiers

72. An important consideration in the design of tuned radio-frequency amplifiers is the increase of selectivity obtained by using two or more stages in cascade. Curve (i) of fig. 37 shows the selectivity of a single tuned anode stage, and corresponds with curve (i) of fig. 27; instead of being plotted as gain against frequency, however, it is shown as "decibels below response at resonance" against "kilocycles off resonance," and it is seen that for equal input voltages a signal 66 kc/s off resonance is attenuated by 6 db. With two such stages, and no regenerative inter-stage coupling, the same reduction is obtained when the interference is 42 kc/s off resonance (curve ii), and when 32 kc/s off resonance in the case of three similar stages (curve iii). If the gain at resonant frequency is A , the overall gain of n stages is theoretically equal to A^n , although in practice it is usually less owing to the effect of the input admittance of each valve upon the gain of the preceding stage. It is usual to provide some means of reducing the amplification in order that when the initial field strength is high, the detector valve may not be overloaded; if the original selectivity is to be maintained this control must operate in some manner other than the introduction of damping into the oscillatory circuits. When screen-grid valves are used, the gain is practically equal to $R_d g_m$, and g_m varies with the mean grid bias to a

greater extent than with triodes. It is therefore possible to control the amplifier gain by applying negative grid bias, and this arrangement is fairly satisfactory in W/T amplifiers, although for R/T reception, it is not entirely so. Alternatively, a variation of screen potential may be employed for gain control. The disadvantages of both these methods, when used in R/T receivers, are dealt with in Chapter XII.

73. The circuit diagram of a typical receiver designed in accordance with the foregoing principles, is given in fig. 38. Two stages of radio-frequency amplification are employed in order that considerable selectivity may be achieved. The inductances L_a , L_b and L_c are of equal value, and are tuned by the condensers C_a , C_b , C_c , which are mounted on a common axis and are capable of simultaneous variation by a single tuning control. Such condensers are said to be "ganged". Controllable reaction is obtained by the variable condenser C_r , the maximum capacitance of which may be only a few micro-microfarads. This is virtually in parallel with

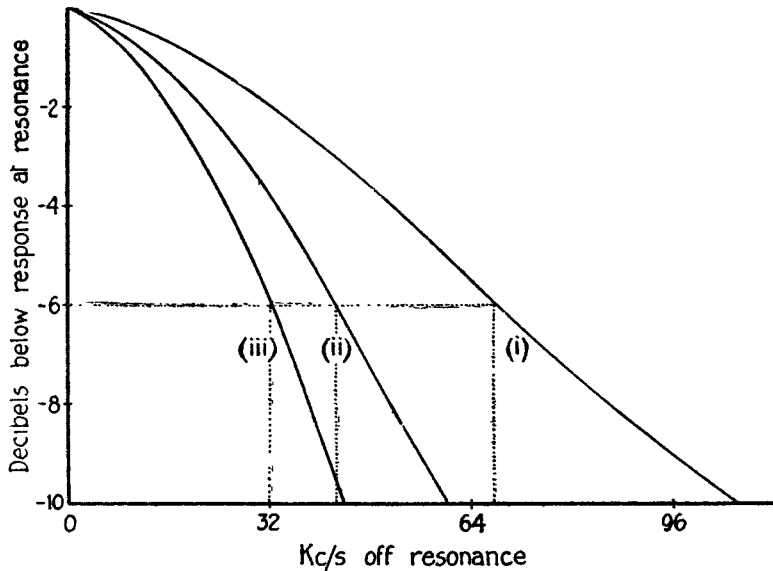


FIG. 37, CHAP. XI.—Selectivity of several R.F. stages in cascade.

the grid-anode capacitance of the second R.F. amplifying valve. The third valve operates as a cumulative grid rectifier, and is transformer-coupled to a stage of audio-frequency amplification, a similar stage being interposed between this and the final valve, which operates as a power amplifier. The gain is controlled by the potentiometer R_s , by which the screen potential of the amplifying valves may be adjusted between limits of about 40 to 80 volts.

Self oscillation in multi-stage amplifiers

74. In addition to the effects caused by grid-anode capacitance and fortuitous inductive coupling between stages, one important form of inter-stage coupling is that caused by the employment of a single H.T. supply for all valves in the receiver. The internal resistance of a new 120-volt battery may be only about 20 ohms, but during use this increases and may reach several hundred ohms before the battery is discarded. This resistance is common to all the anode circuits in the receiver, and may give rise to transfer of energy between them, the result being to cause either positive or negative reaction effects, both of which are undesirable. The frequency of the oscillation caused by positive reaction may be from one to several thousand cycles per second; when of the order of from 4 to 8 cycles per second the resulting noise emitted is often referred to as "motor-boating". It must be realized that the unwanted coupling cannot

CHAPTER XI.—PARA. 75

be reduced to any extent by the use of a condenser in parallel with the H.T. battery, e.g. at 10 cycles per second a $1\mu F$ condenser has a reactance of 1.6×10^4 ohms, which is very large compared with the internal resistance of the battery, and in any event is still common to all circuits of the amplifier. Each A.F. anode circuit is therefore given its own by-pass or decoupling condenser; in fig. 38 these are denoted by C_1 and C_2 . The greater portion of the varying component of anode current is confined to the capacitive path by the insertion of decoupling resistances R_1 , R_2 . Each decoupling resistance may be from 500 to several thousand ohms, the larger values being used when a certain fall of voltage is permissible, e.g. in the detector and early A.F. stages, while the decoupling condensers are usually of from .5 to 2 microfarad. Radio-frequency stages are usually sufficiently decoupled by the use of a condenser only, provided it is inserted immediately adjacent to the anode load inductance as shown in fig. 38. Decoupling is also necessary in the screen circuits of the R.F. stages when these are fed from a single potentiometer, owing to the large common resistance which then exists. In fig. 38 these resistances and condensers are indicated as R_3 , R_4 , C_3 , C_4 . When the H.T. supply is drawn from either D.C. or A.C. mains, the effective internal resistance of the supply device may be many hundred ohms, and decoupling is an essential feature of such receivers, while if the grid bias is provided by any device offering a high impedance, decoupling is also necessary in the grid circuits.

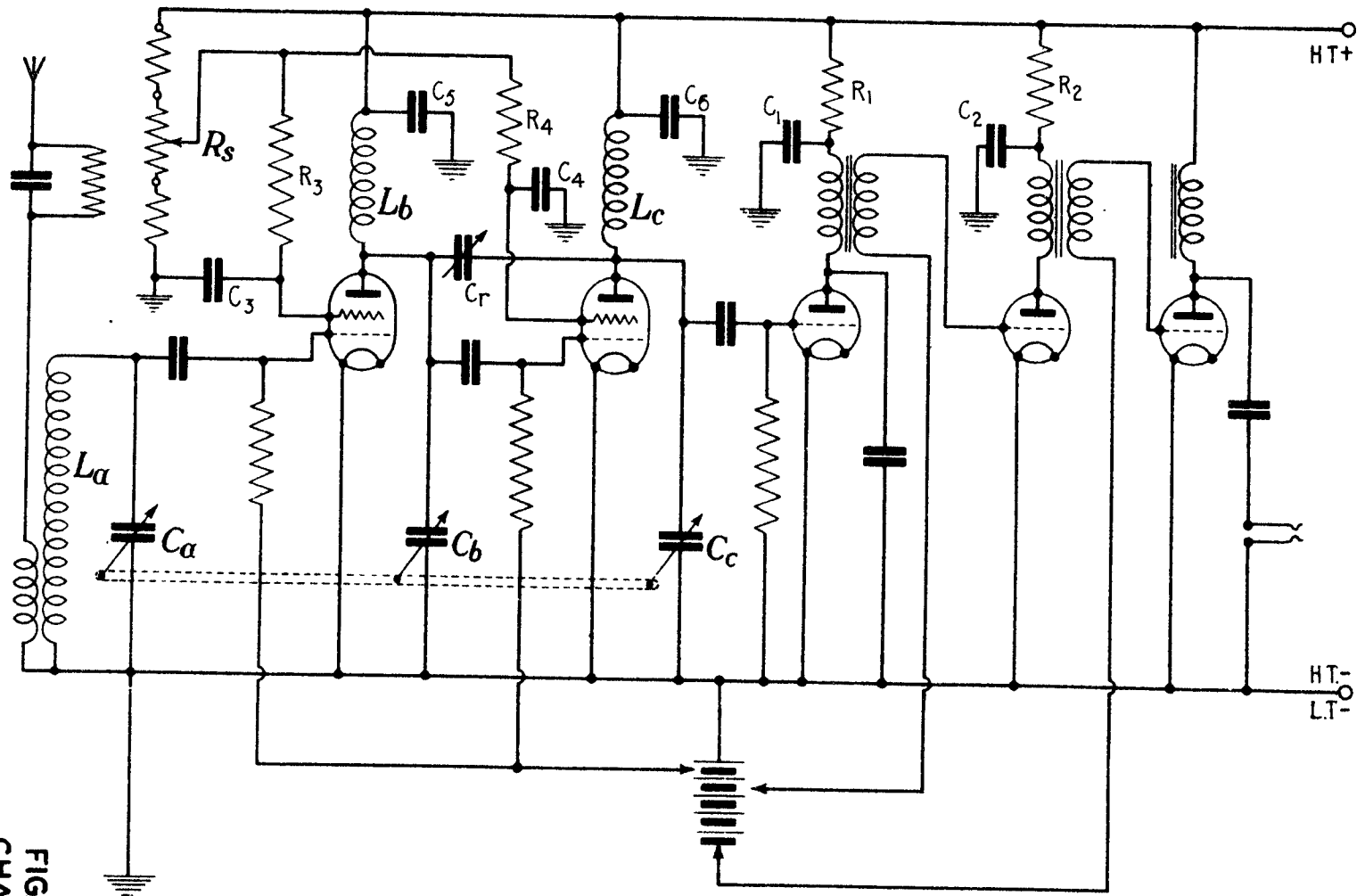
Methods of tuning

75. (i) In the circuit diagrams illustrating the various types of radio-frequency amplifier, the various circuits have been drawn on the assumption that for a given frequency band—usually referred to as the tuning range of a particular inductance—the latter is of fixed value, the frequency adjustment being performed by means of a variable condenser. Where an amplifier is required to cover a wide frequency range, e.g. from 3 Mc/s to 100 kc/s, it is necessary to use a number of inter-changeable coils each of which, in conjunction with the tuning condenser, will cover a definite portion of the total frequency band. For example, let us consider a tuned-anode amplifier in which the capacitance of the tuning condenser (including all distributed capacitance in parallel therewith) may be varied from .00005 to .0005 μF . An 80 μH coil will then cover a frequency range of from 2,500 to 796 kc/s, a 160 μH coil from 1,780 to 560 kc/s, a 320 μH coil from 1,250 to 398 kc/s and a 640 μH coil from 890 to 280 kc/s. To cover the frequency range from 2,500 to 280 kc/s therefore, we actually require only 2 coils, of 80 μH and 640 μH respectively. The provision of an intermediate coil of 160 μH will however allow the tuning condenser to be used near its mean value over a greater portion of its range and will tend to give a greater approach to uniform selectivity.

(ii) An alternative method of tuning is to employ a fixed capacitance for each frequency band and to adjust the value of the inductance in order to tune to the desired frequency. The most convenient method of attaining this end is by varying the magnetic properties of the coil, and is usually referred to as permeability tuning. Although the practical application was attempted many years ago really satisfactory methods of permeability tuning have only been

developed in recent years. The chief object in view is to maintain the ratio $\frac{L}{R}$ at a high and appreciably constant value over the whole tuning range of any particular coil. Since a fixed capacitance is associated with the latter, this also implies a nearly constant stage gain and an approach to uniform selectivity. The practical realization of this ideal depends upon the introduction of a core of comparatively low reluctance, which is composed of iron dust bound with a phenolic compound. The introduction of iron in any shape or form in radio-frequency tuning circuits is, however, fraught with considerable difficulty. The basic idea is, of course, to alter the resonant frequency of the circuit, but the earlier attempts were found to introduce such heavy losses that the circuit became practically aperiodic. In the modern method of permeability tuning the core is made of iron dust of a high purity. It must be very finely divided

FIG. 38
CHAP. XI



TYPICAL RECEIVER

and is generally produced by the chemical reduction of iron oxide. The diameter of the particles is only of the order of 5×10^{-6} inch. Each individual particle must be separately insulated from all others, and this necessitates the use of a special varnish which has high viscosity and yet is capable of forming a very thin film under pressure of the order of 20 tons per sq. in. The particles are afterwards incorporated with phenolic resin to form a substance which resembles solid iron in appearance and actually contains more than 90 per cent. of iron, its specific gravity being of the order of 5, as against 7.8 for solid iron.

Amplifier noise

76. In any form of receiving circuit it will be found that some sound output exists under all circumstances, in addition to that caused by the desired signal or by electrical interference. This is often termed "amplifier noise". A portion of the noise is the result of such obvious causes as induction from power circuits, run-down accumulators, exhausted H.T. and grid bias batteries, faulty contacts, defective components and mechanical vibration of valve electrodes. The remedy in most of these instances is in the hands of the operator, and suitable steps for the elimination of electrical interference have been suggested in the previous chapter. Some valves are more "microphonic" than others, filament tension being one factor in this respect. Specially sprung valve holders are a palliative usually adopted. When all these causes are non-existent, there is always a certain amount of noise which cannot be eliminated because it arises from the actual mechanism of the valve itself. There are three distinct types of spontaneous fluctuation of anode current in any amplifier, which are known as (i) the "shot" effect, (ii) the "flicker" effect, and (iii) the temperature effect. The existence of these effects imposes a limit to the amplification of very small voltages.

The shot effect results from the fact that the anode current is actually a stream of discrete particles rather than the continuous flow generally envisaged. The presence of a space charge tends to smooth out any irregularity in the rate at which electrons arrive at the anode, and it is therefore necessary to provide ample emission from the cathode if the shot effect is to be maintained at a low level.

The flicker effect is probably chiefly caused by the occasional emission of positive ions from the cathode, and to ionisation of residual gas molecules. Whereas the shot effect occurs to the same degree irrespective of the operating frequency, the flicker effect is of greatest importance in audio-frequency amplification, particularly below 1 kc/s. Improvements in the cathode and vacuum may eventually reduce the flicker effect to a negligible amount.

The temperature effect is at once the best understood in theory and the most serious in practice. It is due to the random motion of the electrons within any conductor as briefly described in Chapter I. For engineering purposes it is usual to say that an electron current flows when a large number of electrons are simultaneously moving in the same direction, but in reality every movement of an electron constitutes a current. Although in the absence of an applied E.M.F. the average current is zero, the R.M.S. current may be comparatively large, because even though the average velocity along the conductor is zero, the electrons themselves move about at random in the conductor with a velocity of about 12×10^6 cm. per second. It may also be noted that various electrons or groups of electrons set up voltages at one frequency and other groups at different frequencies, hence the total energy is distributed throughout the frequency spectrum. If the conductor is a portion of the input circuit of an amplifier, the voltage set up in the anode circuit will depend chiefly upon its selectivity. As an example of the magnitude of this effect, it has been calculated that if the amplifier gives effective amplification over a 5 kc/s band, and the input resistance is a .5 megohm grid leak, the noise level produced is equivalent to that produced by an input voltage of 6.4 micro-volt. The total amplifier noise determines the minimum voltage which can be usefully amplified. In the case just cited it is obvious that no advantage will be obtained by using this amplifier to increase the strength of a signal which is initially less than 6.4 micro-volt, for the noise and the signal will be amplified to the same degree and there will be no increase in the ratio of signal to noise.

CHAPTER XI.—PARAS. 77-78

Amplification of high and very high frequencies

77. The gain of a stage of amplification is chiefly dependent upon the possibility of designing an anode circuit impedance comparable with or greater than the anode A.C. resistance of the amplifying valve, and also upon the input admittance, for if the latter is low it has been shown that the input voltage may be reduced to such an extent that the overall gain is actually negative, i.e. a loss. The amplification obtained at high frequencies (with conventional valves and circuits) is always very low, because the input conductance of the succeeding valve shunts the output circuit in the above manner. The input resistance of a cumulative grid rectifier following a R.F. stage may be only 5,000 ohms, and no matter how large the dynamic resistance of the preceding output circuit may be, the effect is to reduce the anode load of the amplifying valve to less than the input resistance of the detector. Little advantage is to be gained by the use of screen-grid valves or R.F. pentodes in such circumstances, as is shown below.

A triode, having an r_a of 20,000 and an amplification factor of 20, works into a resistive load of 5,000 ohms. Find the V.A.F.

$$\begin{aligned}\text{V.A.F.} &= \frac{R_a}{r_a + R_a} \mu \\ &= \frac{5,000}{25,000} \times 20 \\ &= 4.\end{aligned}$$

If the above valve is replaced by a screen-grid valve, $r_a = 300,000$, $\mu = 300$, the V.A.F. is

$$\begin{aligned}\frac{v_a}{v_g} &= \frac{5,000}{305,000} \times 300 \\ &= \frac{5 \times 300}{305} \\ &= 4.9.\end{aligned}$$

The overall gain will be higher than appears from the latter figure because the input admittance of the S.G. valve will be greater than that of the triode and so will not damp the input circuit to the same extent. This is not a serious disadvantage of the triode, however, because reaction may be used to counteract the damping due to the valve. The chief disadvantage of reaction is the difficulty of very smooth control at such high frequencies. The reduction of inter-electrode capacitance in modern R.F. pentodes has made R.F. amplification with conventional circuits feasible up to about 10 Mc/s.

78. By means of special circuits and valves (e.g. the magnetron, chapter IX) it is possible to produce electro-magnetic waves of exceedingly high frequency; in this connection it will be recalled that about 300 Mc/s is the limit above which ordinary reaction methods fail to generate oscillations because the duration of a single cycle is comparable with the average time taken by the individual electrons to travel from cathode to anode. Ordinary valves also fail to function as amplifiers and rectifiers at such high frequencies for the same reason, while the following factors are contributory to this failure :—

(i) The inter-electrode capacitance of the valve is so large that the ratio $\frac{C}{L}$ is too great to achieve any considerable dynamic resistance.

(ii) The inductance of the connections between the external contacts and the actual electrodes is considerable, and results in an appreciable reactive drop.

(iii) The inductance referred to in (ii) acts in conjunction with the inter-electrode capacitance to form undesired tuned rejector circuits.

(iv) The R.F. resistance of the valve is increased by the presence of these inductive and capacitive paths.

Acorn valves

79. In any given valve design, if all the physical dimensions are reduced in a common ratio, there will be no change in mutual conductance, anode current and amplification factor, for given operating potentials. On the other hand the inter-electrode capacitance, lead-in inductance and time of electron transit are all reduced in direct proportion to the reduction ratio. This principle may be expressed by the statement that for optimum performance the linear dimensions of all apparatus should be proportional to the wavelength. At low frequencies, however, adherence to this principle would lead to inconvenient dimensions with little improvement in

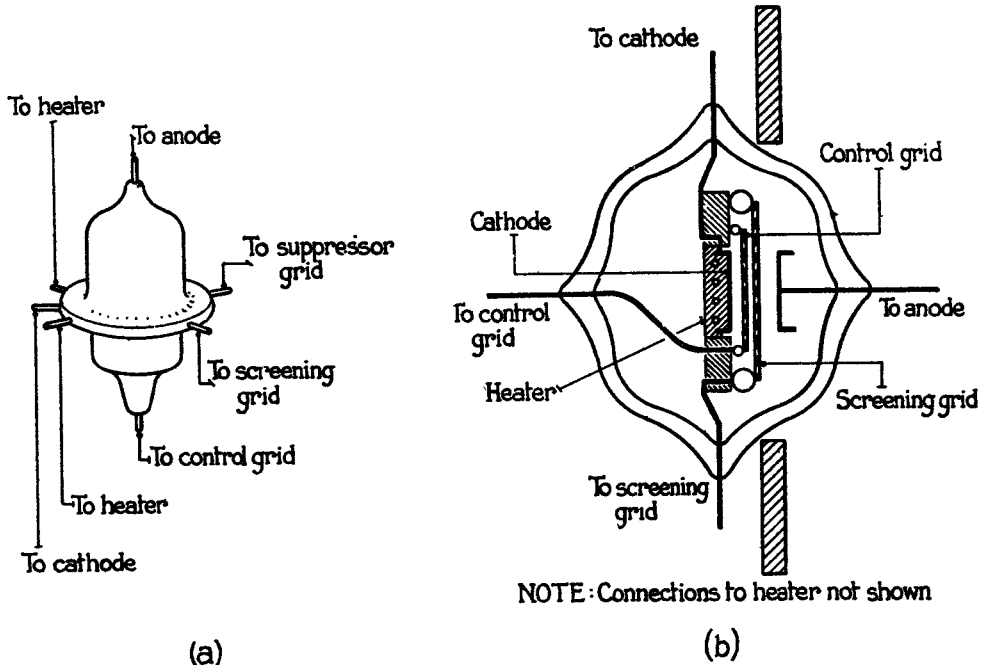


FIG. 39, CHAP. XI.—(a) Acorn R.F. pentode, actual size.

(b) Arrangements of electrodes and connecting leads in acorn tetrode.

performance, but it has been used to produce valves giving appreciable amplification at frequencies up to 600 Mc/s. As the lower limit of useful application of say a standard R.F. pentode is about 5 metres, approximately the same performance can be obtained at .5 metre if all the dimensions of the valve are reduced to one-tenth the normal. Valves produced according to this theory are often referred to as "acorn" valves, the term giving a good idea of the actual size of the valve. Fig. 39a is a perspective drawing of an R.F. pentode of this type, actual size, and also shows the manner in which connections are made to the electrodes, while the latter point is further illustrated in the enlarged sectional drawing of the screen-grid valve of this type (fig. 39b). It will be noted that the electrodes are so arranged and supported that no "pinch" is necessary. The cathode is indirectly heated, but as the current consumption is quite low (about equal to that of a small power valve) the heating current may be derived from a secondary battery. No valve of this type has so far been standardized for service use. Different makers use different heater voltages and slightly different arrangements of lead-in pins. If valves of this kind are used for experimental purposes low-loss valve holders of porcelain or other efficient dielectric should be used, but on no account should connections be soldered directly to the valve, as the glass may soften and allow air to enter the interior.

CHAPTER XI.—PARAS. 80-82

80. Conventional circuits may be employed for R.F. amplification, detection and the production of oscillation up to about 600 Mc/s, provided great care is taken in the arrangement of circuits, the use of shortest possible connecting leads, etc. Where by-pass condensers are used, these should be of quite small capacitance, and the inductance of their connectors reduced to the lowest possible limit. The principle of reduction of linear dimensions should be applied as far as possible to circuit components such as coils and condensers. For example, a coil of 5 turns of 25 s.w.g. copper (diameter, .02 inch) .125 inch in length and of the same diameter, has an inductance of .0587 μH , and a resistance of .37 ohms, at a frequency of 600 Mc/s. The capacitance required to tune this to the stated frequency is 1.2 $\mu\mu F$ and the dynamic resistance of such a circuit is 132,000 ohms, which is higher than is ordinarily obtainable at medium frequencies. Its magnification is of the order of 200. These theoretical values are undoubtedly much higher than are practically attainable. For example, the eddy current loss due to the proximity of both metallic and dielectric materials must be very serious and its magnitude incalculable. Nevertheless it is possible to achieve dynamic resistances and magnifications comparable with those attainable at medium frequencies.

THE SUPER-HETERODYNE RECEIVER

Principle of operation

81. The principle of heterodyne reception of C.W. signals is discussed in Chapter IX, and the supersonic heterodyne or super-heterodyne receiver operates upon similar principles. If a C.W. signal of 1,000 kc/s is to be received by the ordinary heterodyne method, the local oscillator may be set to say 1,001 or 999 kc/s, so that heterodyne beats occur 1,000 times per second. After rectification the output of the detector valve possesses (amongst many others) a component having a frequency of 1,000 cycles per second, which is then amplified as necessary and so caused to operate telephone receivers or loudspeaker at the desired power level. If the frequency of the local oscillator differs from that of the incoming signal by more than 1 kc/s the telephonic response is of higher pitch, and with a sufficiently large frequency difference will be above the limit of audibility as shown in fig. 14, Chapter X. It is important to observe, however, that although no audible sound is produced, the anode current of the detector valve still contains a component having a frequency equal to the number of beats per second. Thus, if the signal frequency is 1,000 kc/s and the local oscillator set to 800 kc/s or 1,200 kc/s, the difference is 200 kc/s and the anode current of the detector valve contains a 200 kc/s component. An output circuit tuned to this frequency will then be set into oscillation and becomes, virtually, the source of a signal having a frequency which is entirely controlled by the adjustment of the local oscillator; such a combination of rectifier and local oscillator is said to function as a frequency-changer. It is also important to note that although the local oscillation is continuously maintained, the new frequency is only produced when a signal is received. To avoid confusion the oscillator used to produce this new radio-frequency will be referred to as the R.F. oscillator.

82. The frequency to which the incoming signal is transferred is called the intermediate or supersonic frequency and is often abbreviated to S.F. A super-heterodyne receiver therefore consists of the following :—

- (i) Radio (signal) frequency circuits, preceding the frequency-changing valve, or valves.
- (ii) The R.F. oscillator circuits.
- (iii) Intermediate or supersonic frequency amplifier circuits.
- (iv) Intermediate or supersonic frequency rectifying valve; this is also referred to as the second detector.
- (v) Audio-frequency amplifying circuits.

In its original form some means of rectification was invariably incorporated in the frequency-changer, and this stage is also sometimes referred to as the first detector. The modern frequency-changing valve performs a somewhat complicated operation upon the signal and the local oscillation, and is therefore sometimes called a "mixing" valve. By whatever name it may be known, it must be fully realized that its function is something more than a mere algebraic addition of the two voltages or currents.

83. If the original signal is of modulated wave-form, the output of the frequency-changer will also be modulated and the output of the second detector will have an audio-frequency wave-form similar to the modulation envelope of the original signal. During the reception of a C.W. signal, the frequency-changer will give an output of unvarying amplitude, and a second heterodyne oscillator is necessary in order to obtain an audio-frequency output from the second detector. This may be an entirely separate oscillator, or alternatively, one stage of the S.F. amplifier may be operated in a self-oscillatory state. The former is generally to be preferred

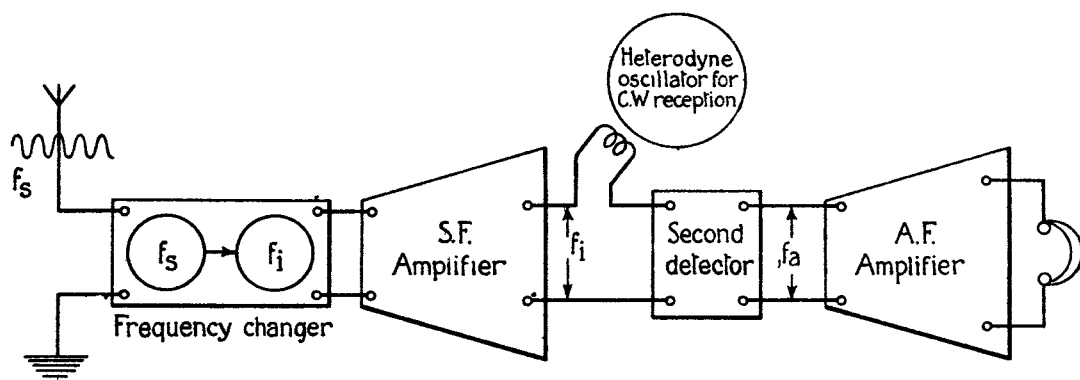


FIG. 40, CHAP. XI.—Schematic diagram of super-heterodyne receiver.

for several reasons, among which are, first, the amplitude of the local oscillation is more easily controlled; second, the overall gain of the S.F. amplifier is greater, because if one stage is operated in a stage of self-oscillation it cannot contribute appreciably to the overall gain. The complete super-heterodyne receiver is shown schematically in fig. 40. For brevity, the oscillator required for C.W. reception will be referred to as the C.W. oscillator.

84. The advantages of the super-heterodyne receiver over the tuned radio-frequency amplifier are :—

(i) However high the signal frequency may be, the actual amplification is performed at the intermediate frequency, which is very much lower. The difficulties associated with the amplification of high radio frequencies are therefore overcome. This statement is not entirely applicable to the first S.F. amplifier of the double super-heterodyne receiver, which is dealt with later.

(ii) By suitable design and adjustment of the frequency-changer the same intermediate frequency is always obtained, and the S.F. amplifier may therefore be designed to amplify only a band extending from five to ten kc/s each side of the nominal S.F. Once they have been properly set up, these stages require no tuning adjustment whatever.

(iii) The overall selectivity is greater than that of the tuned radio-frequency amplifier.

The various forms of frequency-changer will be described briefly, before discussing the factors governing the choice of an intermediate frequency and the selectivity.

CHAPTER XI—PARAS. 85-86

Single-triode frequency-changer

85. This is the simplest type, and the action of all others is easily followed if its operation is understood. The circuit diagram is given in fig. 41. The triode maintains the aerial circuit in oscillation by means of the reaction coupling between the coils L_1 L_2 . Suppose a C.W. signal of 1,000 kc/s is to be received, and the S.F. amplifier stages to be tuned to 200 kc/s. The triode must then maintain the aerial circuit in oscillation at either $1,000 + 200 = 1,200$ kc/s or $1,000 - 200 = 800$ kc/s, and the circuit is, therefore, out of resonance with the incoming signal. This must cause a slight reduction in signal strength, which however is negligible at such high

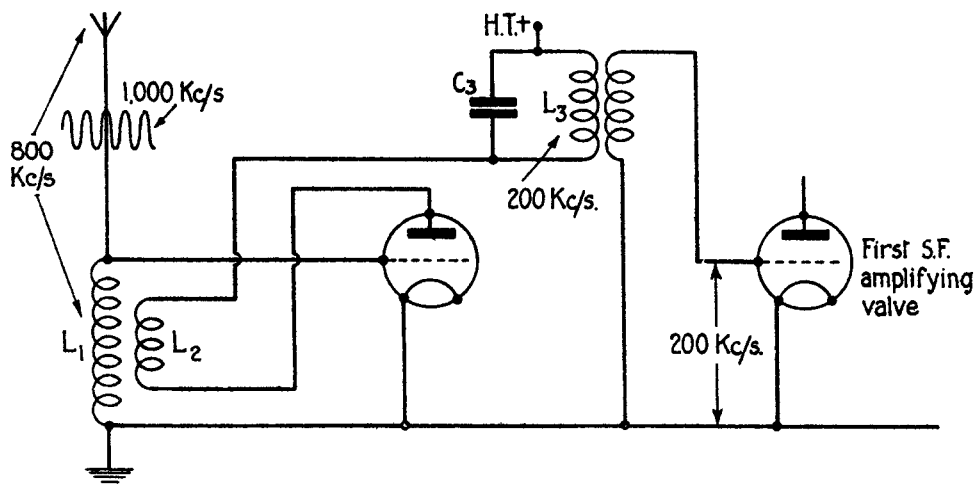


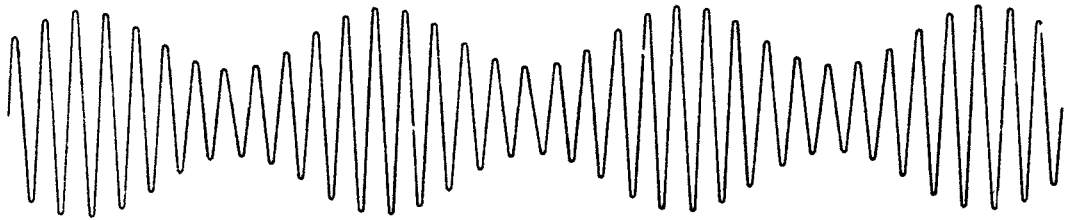
FIG. 41, CHAP. XI.—Single-triode frequency-changer.

frequencies. The amplitude of the oscillatory voltage between grid and filament, in the absence of any signal, must be fairly large, i.e. of the order of ten volts; the resulting heterodyne beats will resemble those of fig. 42a. The $I_a - V_g$ curve of a suitable valve is shown in fig. 42b, with this voltage applied to the grid. The negative edge of the beat envelope has no effect upon the anode current, because it is located beyond the cut-off point on the grid voltage base-line, and the anode current variations may be considered to be due entirely to the positive edge. In this respect the valve functions in a manner somewhat similar to a lower-anode-bend rectifier, hence the term "first detector." The anode current varies in a complex manner, possessing components of 1,000 kc/s and 800 (or 1,200) kc/s, as well as $1,000 + 800$ (or $1,000 + 1,200$) kc/s. The important component however is that caused by the beat envelope, a.b.c.d.e; this corresponds to an anode current component of $1,000 - 800$ (or $1,200 - 1,000$) kc/s, i.e. having a frequency of 200 kc/s. The output circuit (L_3 C_3 , fig. 41) is tuned to this frequency, and will set up an appreciable oscillatory current in the tuned circuit. The resulting voltage across the coil L_3 will therefore be applied to the S.F. amplifier, and will eventually affect the telephones in the usual manner.

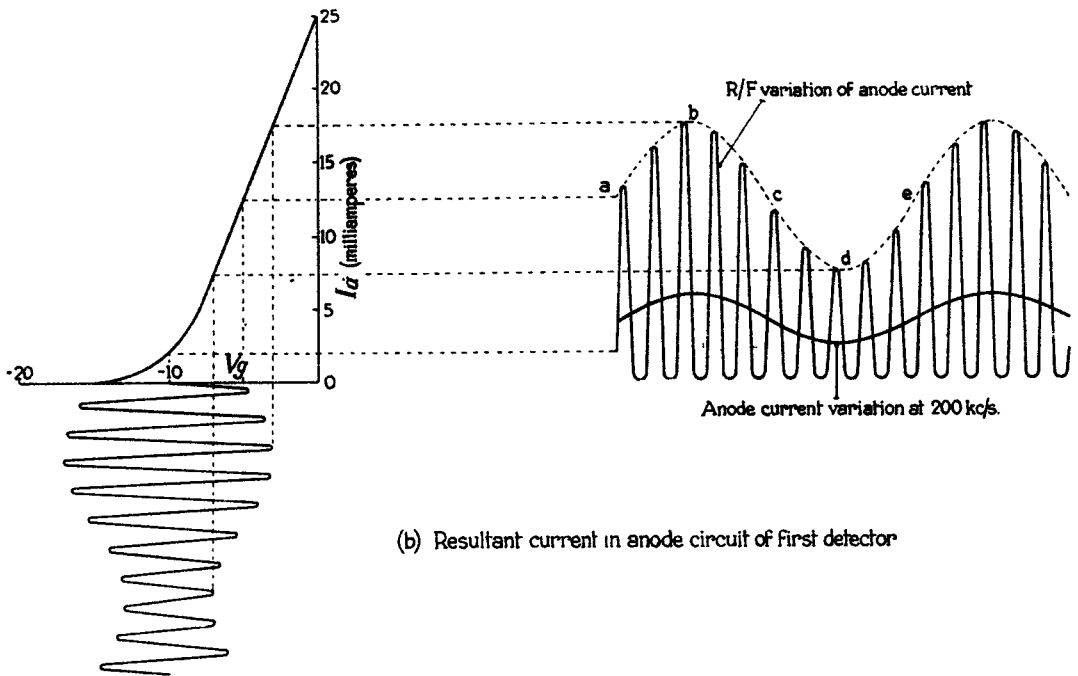
Conversion conductance

86. Although the action is complicated by the rectification process, the action of the frequency-changer is analogous to that of a radio-frequency amplifier, in that a signal voltage $v_s = \mathcal{V}_s \sin \omega_s t$ causes an anode current variation i_a where $i_a = g_c \mathcal{V}_s \sin \omega_i t$, $\frac{\omega_i}{2\pi}$ being the intermediate frequency.

The constant g_c , i.e. the ratio $\frac{\text{intermediate frequency component of anode current}}{\text{signal frequency input voltage}}$ (in the absence



(a) Heterodyne beat between signal and local oscillation



(b) Resultant current in anode circuit of first detector

FIG. 42, CHAP. XI.—Operation of single-triode frequency-changer.

of any anode load), is termed the conversion conductance of the valve. It is analogous to the mutual conductance of the valve when used as an amplifier. When the anode circuit contains an S.F. dynamic load R_d , the relation between i_a and v_s is

$$i_a = \frac{g_c}{1 + \frac{R_d}{r_c}} v_s$$

where r_c is the effective anode A.C. resistance over the operating portion of the $I_a - V_g$ curve ; r_c is sometimes called the conversion resistance of the valve.

87. In fig. 42b, the local oscillation has an amplitude of 5 volts and the amplitude V_s of the signal voltage is 2 volts, the amplitude of the beat voltage envelope varying between 7 and 3 volts. The anode circuit load is assumed to be absent, and the S.F. component of anode

CHAPTER XI.—PARAS. 88-89

current (shown in heavy line) has an amplitude of 1.5 milliamperes. The mutual conductance g_m is 2.5 milliamperes per volt, and the conversion conductance g_c is 1.5 milliamperes $\div$ 2 volts = .75 milliamperes per volt. In this particular instance $g_c = \frac{g_m}{3}$. The conversion conductance of a given valve depends upon the grid bias and the amplitude of the local oscillation as well as upon the shape of the $I_a - V_g$ curve. Under linear rectification conditions, to which fig. 42 approximates, g_c approaches $\frac{g_m}{\pi}$ as a limiting value. With a square law rectifier g_c is rarely greater than $\frac{g_m}{4}$. Owing to the presence of the local oscillation, the conversion resistance of the valve is of the nature of a dynamic resistance rather than of the static nature associated with the usual meaning of the term "A.C. anode resistance." It is possible to measure r_c by plotting an $I_a - V_a$ curve while an alternating voltage is applied between grid and filament, its amplitude being equal to that of the local oscillation under the intended operating conditions; r_c is usually about $2 r_a$.

Conversion factor

88. In the first detector of a super-heterodyne receiver the quantity corresponding to the V.A.F. of an ordinary amplifier is the conversion factor (C.F.), which is defined as the ratio of S.F. voltage across the output impedance to the signal voltage applied to the grid-filament path. From ordinary triode theory it is easily seen that the C.F. of a frequency-changer is given by the equation

$$\text{C.F.} = \frac{g_c r_c R_d}{R_d + r_c}$$

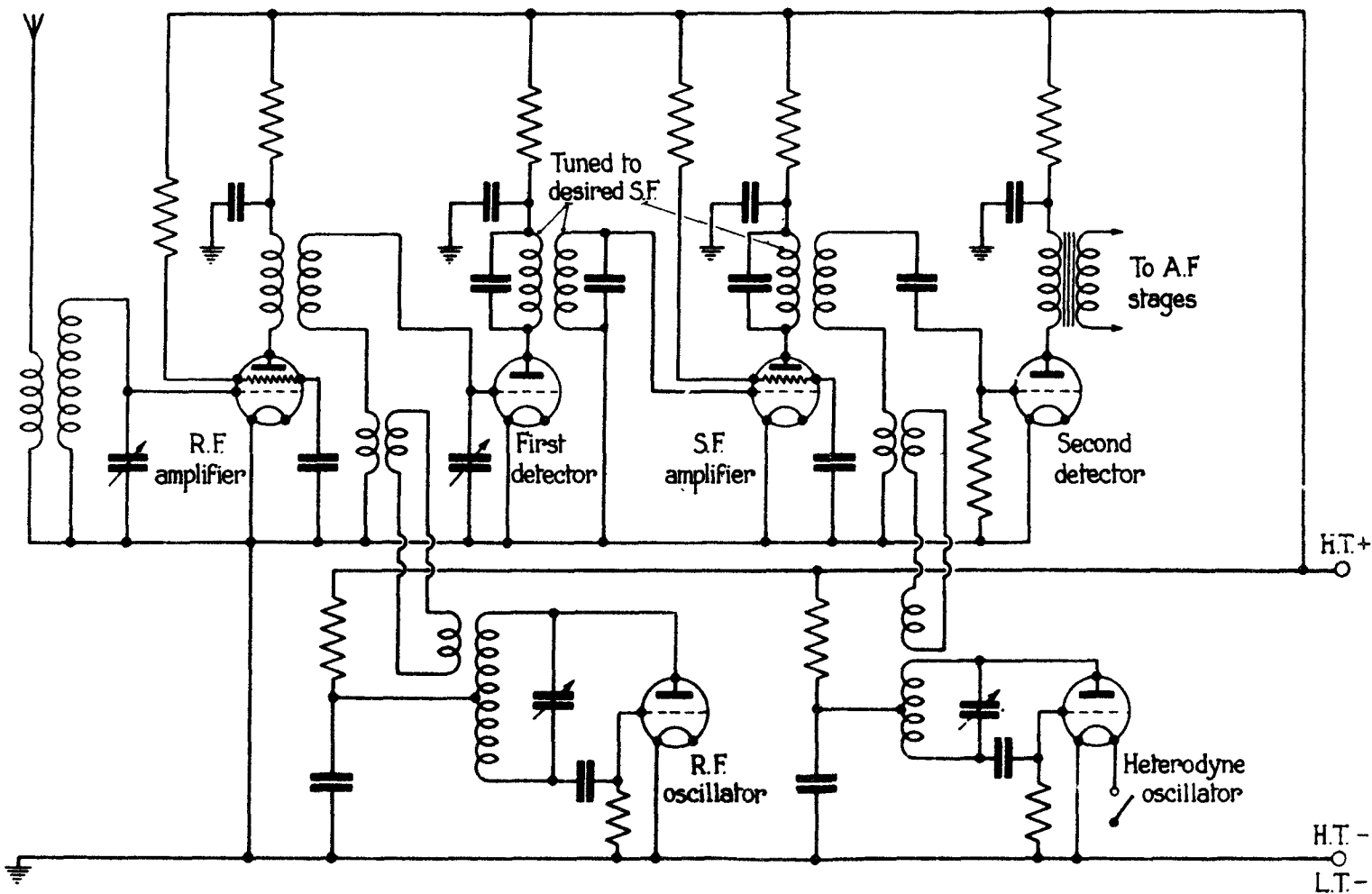
where R_d is the effective dynamic resistance of the load at the supersonic frequency. The product $g_c r_c$ is analogous to $g_m r_a (= \mu)$ in ordinary amplifier calculations. Obviously the conversion gain of various forms of transformer coupling can be dealt with by introducing k and T^2 (coupling factor and inductance ratio) as in paragraphs 48 *et seq.*, the conversion gain being $\frac{v_{g2}}{v_{g1}}$.

Disadvantages of single-triode frequency-changer

89. The single-triode frequency-changer suffers from several disadvantages, the principal being :—

(i) The aerial circuit is in oscillation and therefore radiates energy, causing interference with neighbouring receivers. This can be overcome to some extent by the use of a buffer stage between the aerial and the input to the frequency-changer, which however increases the number of tuning controls.

(ii) The operating conditions, e.g. H.T. voltage and grid bias, must be a compromise between the values best suited for the production of oscillation and those appropriate to a lower anode-bend detector. For maximum conversion conductance, it is desirable that the grid should be biased nearly to cut-off point (but not so negative that oscillations cannot be initiated) and that the grid swing of the local oscillation should not extend into the grid current region. In ordinary oscillators however, the local oscillation builds up to an amplitude which is limited chiefly by damping due to the flow of grid current, so that the conversion conductance must be less than the theoretical maximum. Further, the amplitude of the local oscillation will vary to a considerable extent with the tuning of the input circuit, and the sensitivity of the receiver will vary accordingly. In a certain receiver of this type it was found experimentally that the amplitude of the R.F. oscillation varied between 14 and 30 volts as the aerial tuning was changed from 6 to 3 Mc/s.



SUPER - HETERODYNE WITH TWO - TRIODE FREQUENCY CHANGER

(iii) The receiver as described above is equally effective for two frequencies. Suppose the oscillating aerial circuit to be tuned to 800 kc/s, and the S.F. amplifier to 200 kc/s. Then signals of 600 kc/s and 1,000 kc/s both set up 200,000 heterodyne beats per second and therefore both give rise to an output of the correct intermediate frequency. This is called "image-frequency interference" and can be eliminated by the introduction of suitable tuned circuits between the aerial and the frequency-changer. These are referred to as the pre-selector circuits and may of course be incorporated in a buffer stage.

The two-triode frequency-changer

90. (i) The first of the disadvantages of the single-triode frequency-changer can be overcome by generating the local oscillation in a separate triode oscillator. Fig. 43 gives a possible arrangement in which a signal frequency amplification stage is inserted before the frequency-changer in order to discriminate between the wanted and the image frequency. The R.F. oscillator circuit is coupled to the first detector by means of a link circuit, to reduce so far as is possible any alteration in the tuning of one circuit by the adjustment of the other. The output circuit of the first detector is tuned to the desired intermediate frequency. A screen-grid valve or radio-frequency pentode may be used as the first detector instead of a triode. A modification to this circuit utilizes a buffer or isolator stage between the oscillator and the detector (fig. 44). A screen-grid valve, variable- μ screen-grid valve, or radio-frequency pentode is preferably used in this position, and the buffer then acts as a coupling device for the transfer of energy from the oscillator to the input circuit of the first detector, but not in the opposite direction. Variation of the tuning control of one circuit then has only a negligible effect upon the tuning of the other.

(ii) It will be observed that in both the frequency-changers described above the action is similar to that of the ordinary heterodyne reception of a C.W. signal. The two voltages are first added, giving a heterodyne beat, and the latter is rectified to produce an oscillatory current of the beat frequency. This type of frequency change is sometimes called the additive method.

Multi-electrode frequency-changers

91. The introduction of the mains operated valve led to the trial of a new technique which is referred to as cathode injection. It was first applied to the screen-grid valve but soon led to the development of the triode-pentode, consisting of two entirely separate electrode assemblies in a single envelope, but with a common cathode. The triode section is caused to maintain the R.F. oscillation in a parallel feed circuit, and the incoming signal is applied between the cathode and the control grid of the pentode section. The oscillatory voltage across the grid coil of the triode is virtually in series with the signal voltage and the whole valve functions as an additive frequency-changer, rectification being still required to produce an oscillatory current of intermediate frequency.

The Pentagrid or heptode

92. The cathode injection system has been supplanted by what is called electron coupling in which the action at one electrode is in effect multiplied by that at another. The anode current then contains, amongst other components, a modulation product of the form $A \sin \omega_s t \sin \omega_o t$ where $\frac{\omega_o}{2\pi}$ and $\frac{\omega_s}{2\pi}$ are the frequencies of the R.F. oscillation and incoming signal respectively. Thus both sum ($\omega_s + \omega_o$) and difference ($\omega_s - \omega_o$) components may be expected to exist; the output circuit of the valve is tuned to the latter, and the intermediate frequency amplifier will therefore operate upon the difference frequency. The prototype of one class of electron coupled frequency-changer is the pentagrid or heptode (fig. 45a) which is in effect a triode and screen-grid valve in series. This valve has a cathode, five grids, and an anode, and its I_a-V_a curve exhibits the "negative slope" region peculiar to the screen-grid valve. The grids are referred to by numbers from the cathode outwards, the grid G_1 being used for the purpose of oscillator voltage control, while the grid G_2 acts as the R.F. oscillator anode. As the latter is perforated it allows the electron stream—which pulsates at the oscillator frequency—to penetrate to the outer electrodes. This electron stream, at the point of emergence from the screening grid G_3 ,

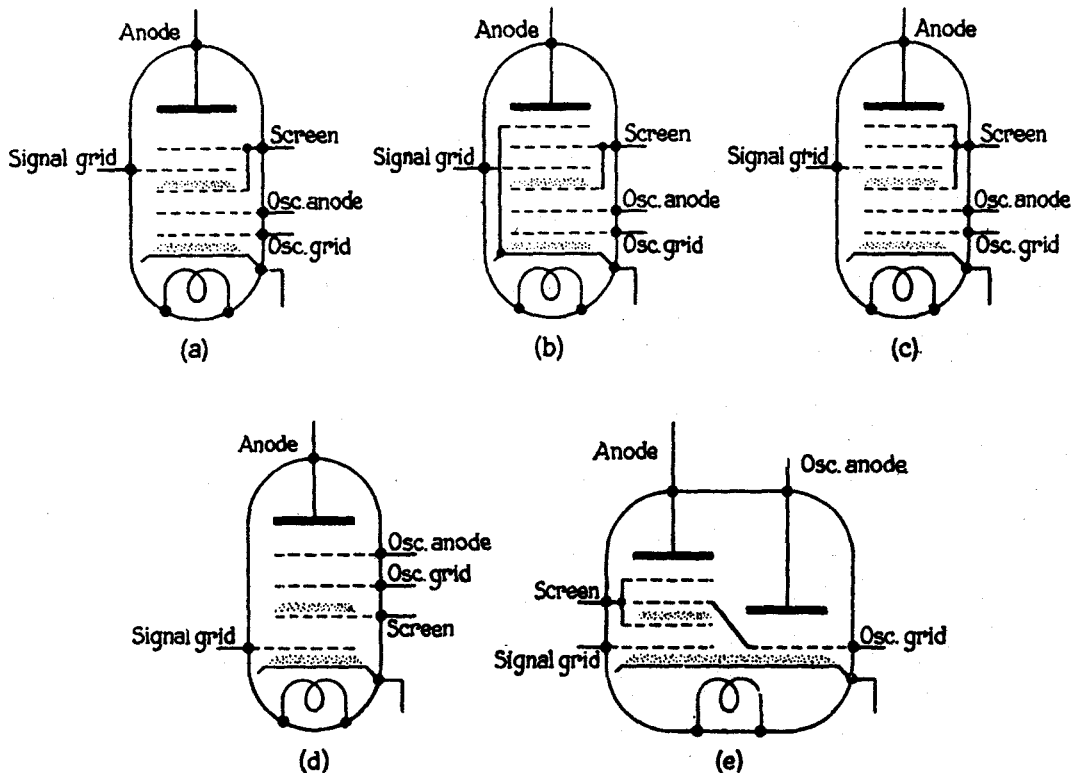


FIG. 45, CHAP. XI.—Frequency-changer valves. (a) pentagrid. (b) suppressor-grid octode. (c) accelerator-grid octode. (d) hexode. (e) triode-hexode.

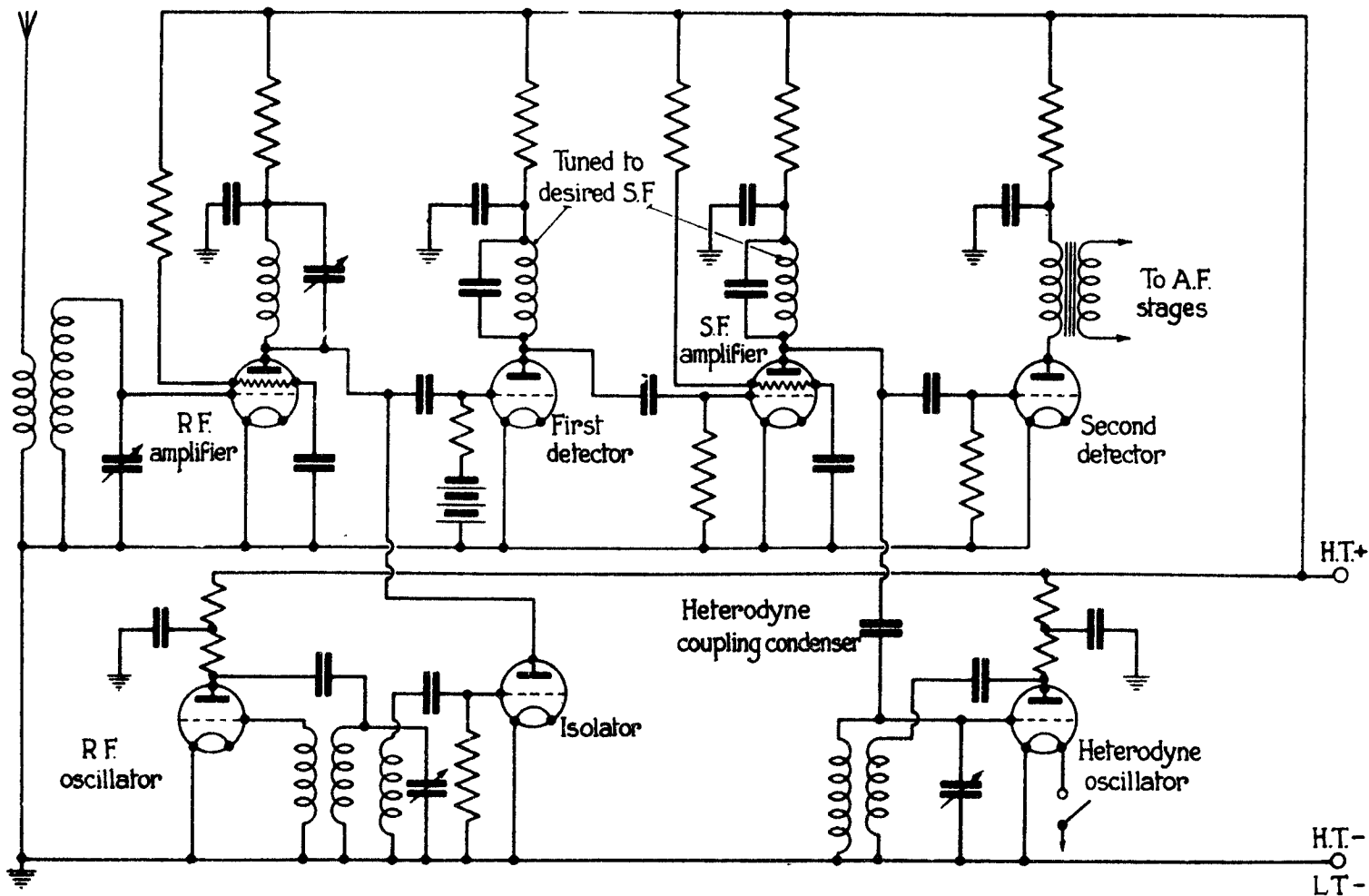
is referred to as the virtual cathode of the tetrode formed by the remaining electrodes. The grid G_4 is the signal voltage control grid, while the grids G_2 and G_3 are connected internally and screen the control grid from the oscillator anode G_1 and from the true anode. The various circuits are connected to the valve as in fig. 46a. If the R.F. oscillator were inoperative, the emission of the virtual cathode would be constant and the mutual conductance $\left(\frac{dI_a}{dV_{g4}} = g_{m4}\right)$

of the tetrode portion also constant. When the R.F. oscillator is generating a frequency $\frac{\omega_o}{2\pi}$ the virtual emission is varied in a sinusoidal manner and the mutual conductance of the tetrode varies correspondingly, becoming in effect

$$g_{m4} = \left(\frac{dI_a}{dV_{g4}}\right) = K g_{m4} \sin \omega_o t$$

For a small sinusoidal change $v_{g4} = \mathcal{V}_{g4} \sin \omega_s t$, in the potential of the signal control grid, the corresponding change in anode current is

$$\begin{aligned} i_a &= g'_{m4} v_{g4} \\ &= K g_{m4} \sin \omega_o t \times \mathcal{V}_{g4} \sin \omega_s t \\ &= \frac{K}{2} g_{m4} \mathcal{V}_{g4} [\cos (\omega_o - \omega_s) t - \cos (\omega_o + \omega_s) t] \\ &= g_c \mathcal{V}_{g4} [\cos (\omega_o - \omega_s) t - \cos (\omega_o + \omega_s) t] \end{aligned}$$



SUPER-HETERODYNE WITH ISOLATOR STAGE

Hence g_c cannot be greater than $\frac{g_{m4}}{2}$ because K cannot exceed unity. The output circuit is tuned to the difference frequency $\frac{\omega_o - \omega_s}{2\pi}$ and supplies excitation to the S.F. amplifier as in the types of frequency-changer previously discussed.

93. It will be observed that the anode current contains a difference component even if g'_{m4} is independent of the anode voltage and the grid bias hence the frequency changing action is not dependent upon the curvature of the static characteristics and the difference frequency is not produced by a rectifying action as in an additive frequency changer. In practice, the curvature of the characteristics may lead to the production of a S.F. by rectification of the beats resulting from the interaction of harmonics of the signal and any interfering frequency which may be present. The conversion gain of the pentagrid is considerably greater than that obtained with the single-triode or two-triode arrangement, but is of the same order as is obtained by employing a tetrode or radio-frequency pentode as the first detector, together with a separate R.F. oscillator.

The octode

94. The octode functions in a similar manner to the pentagrid, the principal difference being the introduction of an additional screening electrode. In the suppressor-grid octode (fig. 45b), this screen is at cathode potential, giving the valve the I_a-V_a curves of a radio-frequency pentode in that the region of negative slope is eliminated. It is therefore, capable (under certain conditions) of giving a larger output than the pentagrid. In certain designs, however, the screening electrodes are all maintained above cathode potential as in fig. 45c, that nearest the anode being called the accelerator grid. The I_a-V_a curves then exhibit a region of negative slope. The chief advantage of the octode is that automatic volume control is more effective, and simpler in application, than in the pentagrid. Its conversion resistance is also higher, giving a slightly higher conversion gain for an equal conversion conductance, while the selectivity of the output circuit is somewhat greater, owing to the smaller damping effect of the valve resistance.

95. The development of other types of frequency-changer, such as the hexode and the triode-hexode, is bound up with the requirements peculiar to broadcast receivers, where the necessity for a single tuning control renders it necessary to gang the R.F. oscillator and signal frequency circuits. In most circumstances, this entails that the R.F. oscillator shall operate on the higher of the two possible frequencies. At high and very high frequencies the input impedance of the frequency-changer is also of importance. The tuned circuit of the R.F. oscillator is coupled to the input circuit by the inter-electrode capacitance, and since the two circuits are not tuned to the same frequency, the input impedance of the valve will have either a positive or a negative resistance component. When the R.F. oscillator frequency is above that of the signal, as in most commercial super-heterodynes, the tuned circuit of the oscillator is, in effect, a capacitive load upon the input circuit and the input resistance of the valve is positive, imposing damping upon the input circuit. As a result of this damping, the effective conversion conductance of the pentagrid falls off at frequencies higher than about 2 Mc/s, and that of the octode above about 5 Mc/s. Where the oscillator is set to the lower frequency, the effective conversion conductance increases at the higher frequencies.

The hexode

96. In both heptode and octode the R.F. oscillator control grid is that nearest the cathode, i.e. G_1 . The prototype of a second type of frequency-changer is the hexode (fig. 45d), which possesses a cathode, four grids and an anode. In this valve the signal voltage is applied between G_1 and the cathode, G_2 is a perforated electrode which is maintained at a positive potential and serves both as a screen and as an accelerating electrode. The grids G_3 and G_4 are the R.F. oscillator grid and anode respectively. In operation, a virtual cathode is formed on the outside

CHAPTER XI.—PARAS. 97-98

of G_2 , and during reception of a signal its emission will pulsate at signal frequency. This emission will be modulated by the R.F. oscillation and on arrival at the anode the electron stream will possess a component of difference frequency as in the pentagrid. The hexode suffers from the disadvantage that automatic gain control is difficult to achieve. This is because the R.F. oscillator derives its electron supply from the virtual cathode. If a heavy negative bias is applied to the signal grid, the virtual cathode becomes non-existent or at any rate of such a low electron density that the R.F. oscillation is no longer maintained. The effect of the inter-electrode capacitance coupling upon the input impedance of the valve is very much less in the hexode than in the pentagrid or octode. In effect, this is due to the relative potentials of the adjacent electrodes. Taking the octode, we see that only a single screen separates the input grid from the R.F. oscillator anode, the potential of the latter being above that of the screen. In the hexode, the signal grid and R.F. oscillator anode are separated by two electrodes, viz. the screen G_3 , at a fairly low positive potential, and the R.F. oscillator grid G_4 , which has a mean potential negative to the cathode. A *prima-facie* examination would therefore lead to the conclusion that the inter-electrode coupling between the tuned oscillator circuit and the input circuit is less than in the octode. Actually, this is found to be the case, although the actual explanation is much more complicated, involving as it does the electro-dynamic effects due to the varying electron density of the virtual cathode.

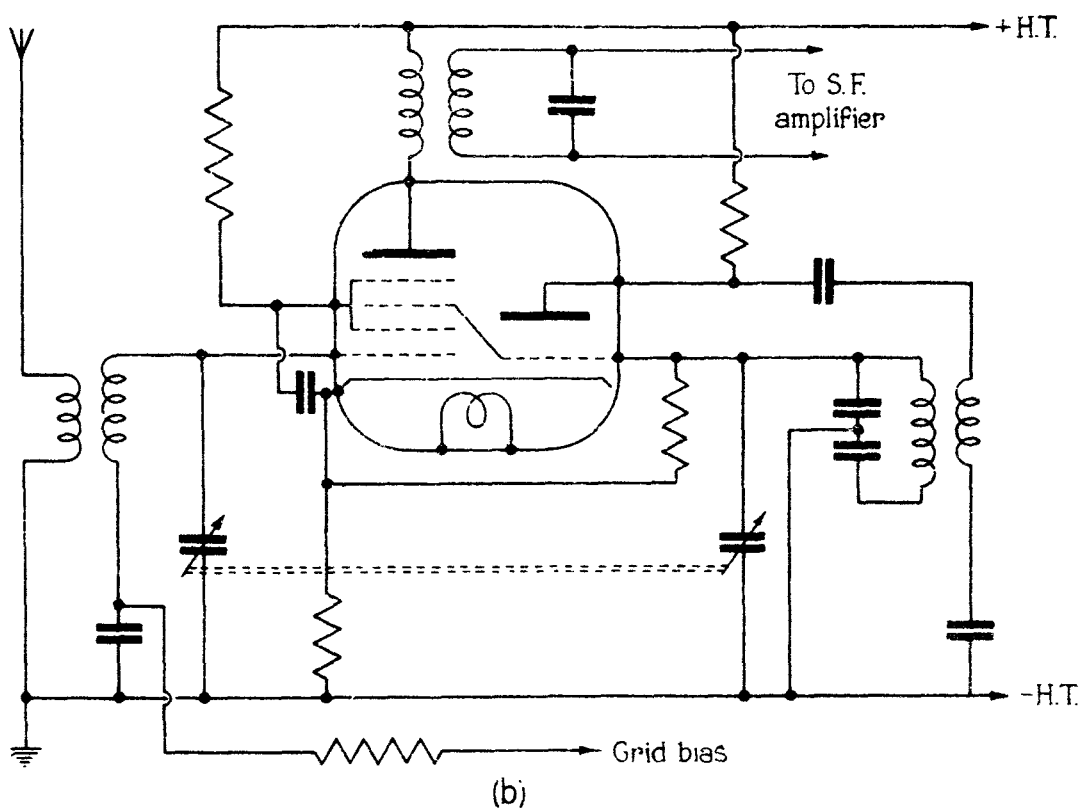
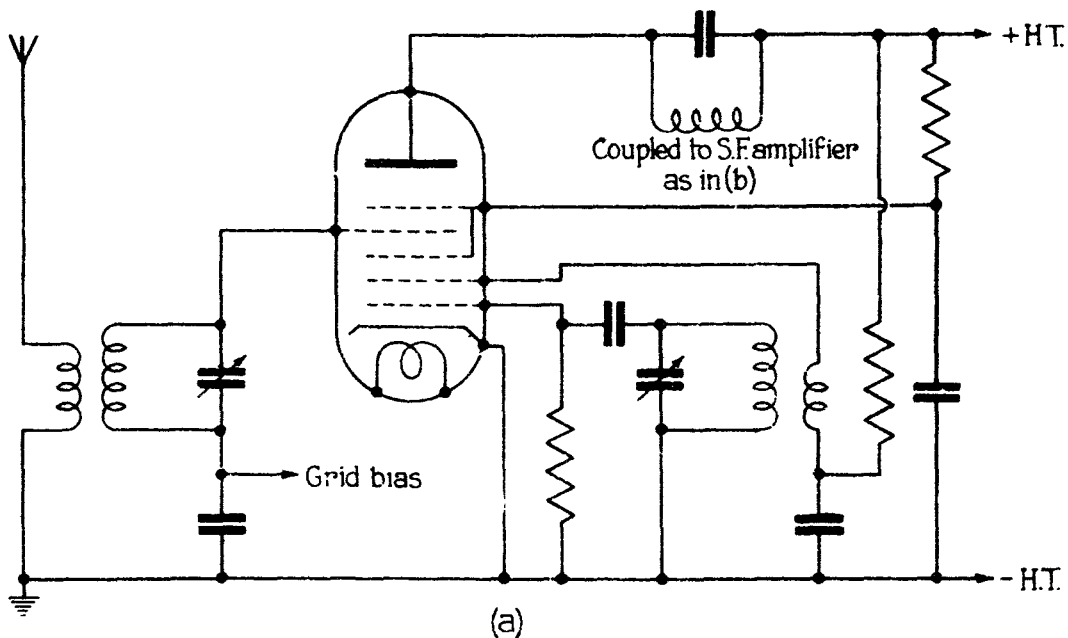
The triode-hexode

97. (i) This valve (fig. 45e) consists of a hexode and a triode in a single envelope, the triode portion operating as an R.F. oscillator. The circuit connections are shown in fig. 46b. In the hexode portion, G_1 is the signal voltage grid, G_2 and G_4 are linked together to form a screening electrode surrounding G_3 which is called the injector grid and is in direct connection with the R.F. oscillator grid. The tuned circuit of the R.F. oscillator is now almost entirely decoupled (both electro-statically and electro-dynamically) from the input circuit, and the input impedance is very little affected by the R.F. oscillator at frequencies below about 100 Mc/s.

(ii) Most of the electron-coupled frequency-changers suffer from a tendency for the R.F. oscillation to pull into step with the signal frequency when the signal grid bias is varied by the automatic gain control. This is again only of importance at high frequencies where the percentage difference between signal and oscillator frequencies is very small. The electro-dynamic properties of the virtual cathode are again involved in a complete explanation, but bearing in mind that the frequency of a valve-maintained oscillation depends to some extent upon the anode A.C. resistance r_a of the valve, the effect is to all intents and purposes due to the variation of the value of r_a with grid bias. It is therefore not encountered in a triode-hexode frequency-changer, in which the electron current of the triode portion is in no way affected by the variation of signal grid bias since this operates only upon the hexode portion.

Choice of R.F. oscillator frequency

98. Except in the case of the single-triode frequency-changer (the input circuit of which is tuned to the signal frequency, plus or minus the predetermined supersonic frequency) the signal frequency circuits are preferably ganged as in the tuned radio-frequency amplifier. Where the receiver is to be operated by unskilled personnel, arrangements are often made to gang the R.F. oscillator tuning also. At first sight this does not appear possible because the oscillator frequency should theoretically differ from the signal frequency by the correct amount, i.e. the supersonic frequency, at all points in the tuning range. It is found, however, that if the minimum capacitance of the R.F. oscillator condenser is set up by means of a small pre-set condenser, in parallel, and its maximum capacitance slightly decreased by the insertion of a fairly large pre-set condenser, in series, the required frequency difference can be maintained at the two ends and at the middle of the tuning range; elsewhere the actual supersonic frequency will not be quite correct. It will be noticed that the effective frequency ratio between the minimum and the maximum condenser setting will be less than that of the variable condenser alone. If the latter will give a 3 : 1 frequency ratio, the presence of the added pre-set condensers will reduce the effective ratio to only about 2.75 : 1. In practice this method of achieving what is



TYPICAL FREQUENCY CHANGERS

called correct "tracking" necessitates that the oscillator frequency shall be higher than that of the signal. As an example, let us suppose the supersonic frequency to be 100 kc/s and the desired frequency range, with a given set of coils, to be 550 to 1,500 kc/s. The oscillator must then tune either from 650 to 1,600 or from 450 to 1,400 kc/s, i.e. over a frequency ratio of 2·4 or 3·1 to 1 respectively. The latter figure entails a variation of capacitance of 1 to 10, e.g. from $\cdot 00005$ to $\cdot 0005 \mu F$. The former requires a variation of only 1 to 5·5, e.g. from $\cdot 00008$ to $\cdot 00044 \mu F$. When this method of ganging is adopted, therefore, the R.F. oscillator invariably operates on the higher frequency.

Selectivity—the S.F. amplifier

99. The high degree of selectivity obtainable in a super-heterodyne receiver has led to its adoption for the reception of frequencies of the order of 200 kc/s and below, in addition to the high and very high frequencies for which it was originally intended. This selectivity is chiefly dependent upon the circuits of the S.F. amplifier. Consider the reception of I.C.W. signals on

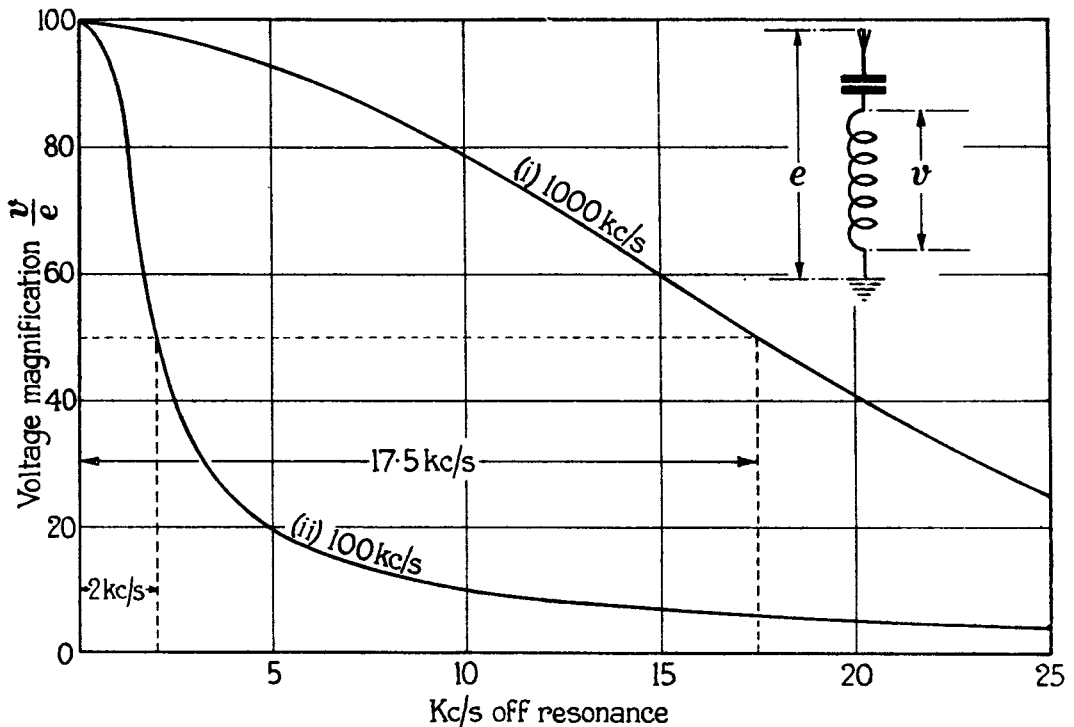


FIG. 47, CHAP. XI.—Comparative selectivity of tuned circuits of equal Q .

1,000 kc/s, first with a simple non-regenerative receiver, the single tuned circuit having a magnification of 100, and the response characteristic given in curve (i) of fig. 47. Assuming equal field strengths, a signal 17·5 kc/s off resonance will give an input voltage to the detector one-half that at resonance and all signals within a band of 35 kc/s may therefore be expected to cause interference. If, however, a similar type of receiver is used on 100 kc/s, and the circuit magnification is the same, namely 100, the response characteristic is that shown in curve (ii), and the same degree of interference is experienced only over a band of 4 kc/s. Now consider the reception of three signals of equal field strength, the frequencies being 990, 1,000 and 1,010 kc/s, by a super-heterodyne having a S.F. of 100 kc/s. The first local oscillator may be adjusted to 1,100 kc/s producing 100,000 beats per second from the 1,000 kc/s signal, and a current of the

CHAPTER XI.—PARA. 100

correct supersonic frequency. This signal will be fully amplified in the S.F. stages. The 990 kc/s signal will give a S.F. current of $1,100 - 990 = 110$ kc/s and the 1,010 kc/s signal a S.F. current of $1,100 - 1,010 = 90$ kc/s. Both the undesired signals are now 10 kc/s, or 10 per cent. out of resonance and will receive negligible amplification. This form of discrimination is referred to as adjacent channel selectivity.

Types of interference peculiar to super-heterodyne

100. (i) Again, if a supersonic frequency of 50 kc/s is chosen, the local oscillator being adjusted to 1,050 kc/s, the interfering signals will give rise to S.F. currents of 60 and 40 kc/s respectively, which are 20 per cent. out of resonance, and will be attenuated to an even greater extent than in the previous case. From this it would appear that a low supersonic frequency should be chosen. If too low, however, another form of interference is liable to occur; for example, suppose an interfering signal to have a frequency of 1,050 kc/s. By the above reasoning it should give no heterodyne beat with the local oscillator and produce no S.F. response, but actually it will form heterodyne beats with the desired oscillation, and these give rise to an S.F. of $1,050 - 1,000 = 50$ kc/s, which is, of course, fully amplified. As this interference only occurs when both transmitters are radiating, it gives rise to a peculiar "mush" which renders reception extremely difficult. When a high supersonic frequency is chosen, this form of interference is reduced, because the frequency causing it is further removed from that of the desired signal and the preliminary tuned circuits, or pre-selector stages, are competent to deal with it; e.g. if the S.F. is 200 kc/s, this form of interference would be caused only by a signal on 1,200 kc/s (or 800 kc/s), 16 per cent. out of resonance. Unless its field strength is considerable, such a signal will be considerably attenuated before it reaches the frequency-changer.

(ii) A form of interference peculiar to the super-heterodyne is that caused by the harmonics of undesired signals. This may be shown briefly as follows:—

	Kc/s.	Kc/s.
Desired signal	1,000	
Supersonic frequency	100	
Local oscillator	1,100	or 900
Undesired signal	1,150	
Supersonic current due to undesired signal	50	or 250

Neither of the latter will give appreciable interference. Now consider the local oscillator and undesired signal to possess a considerable second harmonic content.

	Kc/s.	Kc/s.
Local oscillator, second harmonic	2,200	or 1,800
Undesired signal, second harmonic	2,300	or 2,300
Supersonic current due to undesired signal	100	or 500

With the local oscillator set to a higher frequency than the desired signal, the second harmonic interference will be fully amplified.

(iii) Yet another form of interference results if two signals, the respective frequencies of which are equal to the desired signal plus and minus one-half the S.F., are simultaneously received. The rectification of this combination obviously results in a current of the supersonic frequency. Other complicated reactions often occur, e.g. between the desired signal frequency and harmonics of the supersonic frequency, if the latter are allowed to be transferred to the circuit preceding the first detector. The remedy in this case is adequate screening of signal, R.F. oscillator, and S.F. circuits.

(iv) Signals of the supersonic frequency may be received in the aerial and transferred to the S.F. amplifier by electro-magnetic or electro-static coupling. This interference is reduced by careful screening and by the provision of a high degree of selectivity in the R.F. stages.

101. To sum up, the choice of an intermediate frequency is a compromise between the following factors :—

(a) Given an adequate degree of selectivity in the pre-selector stages, image-frequency interference is reduced as the S.F. is increased. The possibility of interference by two stations, the frequency difference of which is equal to the supersonic frequency, is also reduced by a high S.F.

(b) Low-order harmonics of the S.F. should not fall in the frequency band to be received, e.g. a S.F. of 500 would be most unsuitable for reception of 1,000 kc/s.

(c) It is not possible to receive signal frequencies equal, or nearly equal to the S.F. By suitable switching arrangements the S.F. stages only may be used as an ordinary (tuned radio-frequency) amplifier for this frequency band.

(d) The difficulty of obtaining a high degree of selectivity and amplification increases with an increase of S.F.

The double super-heterodyne

102. (i) A little consideration of the above summary will show that for reception of high radio frequencies of the order of say 20 Mc/s it is difficult or impossible to achieve a high degree of image frequency discrimination by the ordinary super-heterodyne circuit. Thus with a S.F. of 100 kc/s and an oscillator frequency of 20 Mc/s the receiver deals with 20·1 and 19·9 Mc/s

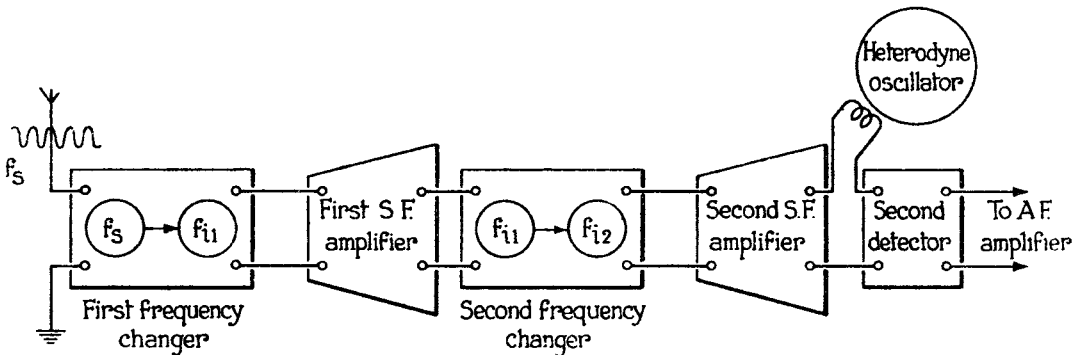


FIG. 48, CHAP. XI.—Super-heterodyne with double frequency change.

equally well, unless the pre-selector stages are capable of giving a very high degree of discrimination against a signal only one per cent. off resonance. If the S.F. is appreciably increased, so that a single pre-selector stage is competent to deal with the image frequency, the adjacent channel selectivity, i.e. that of the intermediate frequency amplifier, will be correspondingly reduced. The difficulty may be overcome by using what is called a double super-heterodyne, the schematic diagram of which is given in fig. 48. This in effect is an ordinary super-heterodyne receiver in which the pre-selector, instead of being a tuned radio-frequency amplifier, is another super-heterodyne receiver with a comparatively high S.F. This S.F. must of course fall within a frequency band of which reception is not required. As an example, suppose it is desired to cover a frequency band of from 20 Mc/s to 150 kc/s except for the region 1,500 to 2,000 kc/s. The first S.F. will therefore naturally fall in the latter band. The lower the frequency the better is the image frequency discrimination. Since however, broadcast transmission may cause direct interference on 1,500 kc/s a little margin must be allowed and 1,700 kc/s is found to be suitable. A single tuned circuit preceding the first frequency-changer will then be found to give adequate discrimination against the image frequency, e.g. for a signal of 20 Mc/s the image frequency

and signal frequency differ by $\frac{2 \times 1.7}{20} \times 100 = 17$ per cent. Even at such a high frequency

CHAPTER XI.—PARAS. 103–104

as this it is possible to provide an aerial circuit magnification sufficient to attenuate the image frequency by about 12 db. At lower signal frequencies the image frequency attenuation is even better. The second S.F. will be dictated by the same considerations as in the single super-heterodyne, viz. it must give adequate discrimination against adjacent channel interference. Let us assume that the chosen frequency is 100 kc/s. The first S.F. amplifier must have sufficient selectivity to act as a pre-selector to the second frequency-changer. Bearing in mind that no matter what the signal frequency may be, the input to the first S.F. amplifier is at 1,700 kc/s, the second R.F. oscillator may be set to either 1,600 or 1,800 kc/s. If the former is chosen the desired signal, 1,700 kc/s, and any interference on 1,500 kc/s, will be fully amplified by the second S.F. amplifier. The first S.F. amplifier must therefore be sufficiently selective to suppress any 1,500 kc/s interference which it may receive from the first frequency-changer valve. The second S.F. amplifier may be provided with variable selectivity for R/T and W/T reception.

(ii) As might be expected, such a receiver must be very carefully designed with a view to reducing image interference and that due to the signal frequency plus or minus one-half the S.F. There are other problems peculiar to the employment of a double frequency change. The number of possible interfering harmonics is reduced by operating all the oscillators (i.e. first R.F., second R.F., and C.W. heterodyne) at the higher of their two possible frequencies, but in practice it is preferable to operate the second oscillator at the lower one. The reason is as follows. When the receiver is operating on 200 kc/s, the first oscillator is on 1,900 kc/s. It is difficult to decouple the various circuits so thoroughly that no beats occur between first and second oscillators. If such a beat does occur with the second R.F. oscillator on 1,800 kc/s, its frequency is $1,900 - 1,800 = 100$ kc/s, which gives rise to a continuous signal in the second S.F. amplifier. With the second oscillator set at 1,600 kc/s this difficulty does not arise.

RADIO-FREQUENCY POWER AMPLIFIERS

103. In most modern transmitting equipment the oscillator operates at a low power level and steps are taken to ensure that the frequency is as nearly constant as possible, having regard to the conditions under which the transmitter is to be employed. The low-power, constant-frequency oscillation is amplified by successive stages until the required power level is reached, the final output circuit being the transmitting aerial with its feeder line and impedance-matching device. In such a transmitter each stage is a power amplifier, and the efficiency of conversion from direct to oscillatory power is of great importance. Where it is possible to provide a sufficiently high anode supply voltage, the power output is limited only by (i) the permissible dissipation, (ii) the filament emission and (iii) the efficiency, of the valve. For high efficiency of power conversion it is necessary to operate in such a manner that anode current flows for not more than one-half the duration of each cycle, and the oscillatory anode-filament P.D. must be nearly equal to the supply voltage. The load impedance of a radio-frequency power amplifier invariably consists of an oscillatory circuit which is tuned to the frequency of the input (grid-filament) voltage.

Angle of current flow

104. It is often convenient to speak of the duration of the anode current impulses in terms of the electrical angle during which anode current flows. In the class B amplifier, the grid is given such negative bias that the anode current, in the absence of an oscillatory grid-filament P.D., is reduced to zero, or very nearly so. The application of an oscillatory P.D. between grid and filament then causes anode current to flow during the positive half-cycles of grid voltage only; the current is therefore said to flow for 180 degrees. These anode current impulses are, to all intents and purposes, half sine waves, and the operating conditions are very similar to those of the high-efficiency oscillator discussed in Chapter IX, except that the grid bias voltage must be maintained at a constant value irrespective of the magnitude of the grid-filament oscillatory voltage. Bias cannot therefore be obtained by means of a condenser and leak resistance. In the class C amplifier, the grid is biased to beyond the point of anode current

cut-off, and anode current flows for less than 180 degrees. It is important to observe that in the class B amplifier, the electrical angle is the same no matter what the amplitude of the grid-filament voltage may be, whereas in the class C amplifier the electrical angle generally varies with the amplitude of the grid-filament voltage.

Limiting conditions

105. Before proceeding to an approximate derivation of the power output and efficiency of these amplifiers, it is necessary to appreciate certain practical limitations which may be imposed. Possibly the most important of these is the filament emission under average working conditions. In some cases this is given by the valve manufacturer, but the following figures may be taken as a working approximation where precise information is not available. As an average during its useful life a pure tungsten filament may be expected to give from 5 to 10 milliamperes for each watt expended in filament heating. The exact figure to be employed depends upon the size—and therefore the cost—of the valve. A small and comparatively inexpensive valve may be run with a higher emission than a large one, for although its life will be shorter the cost of replacement is much less. The following empirical rule may be taken as a rough guide :—

$$I_e = P_f (10 - .005 P_L).$$

where I_e is the filament emission in milliamperes.

P_L is the permissible dissipation in watts.

P_f is the power expended in filament heating, in watts.

According to this rule, a 30-watt valve will give nearly 10 milliamperes per "filament watt", while a 1,000-watt valve will give only 5 milliamperes per "filament watt." The thoriated tungsten filament, such as is used in the V.T. 25 valve, will give a momentary or "flash" emission of from 20 to 40 milliamperes per "filament watt", but in operation the anode current must not be allowed even to approach the flash emission value, otherwise the valve will deteriorate rapidly. It may therefore be assumed that such a filament will give about 15 to 20 milliamperes per "filament watt." Coated filaments are employed in certain power amplifying valves of ratings up to about 50 watts. Their working emission is about 60 to 80 milliamperes per filament watt. A second limitation which arises is the magnitude of the supply voltage, which may be restricted by the compactness of a given installation, and its consequent small margin of insulation resistance, or by the weight and volume of the H.T. generator or other supply device. The maximum supply voltage for a particular valve is generally given on the maker's label, and should not be exceeded.

Class B amplifier

106. The behaviour of a class B amplifier may be approximately determined by a consideration of the current and voltage relations during operation. It is always necessary to make certain assumptions, and the different conclusions reached by various writers are chiefly due to differences in the stipulated conditions. It will be convenient to commence by assuming that the valve possesses ideal characteristics, the notation used being as follows.

I_e = filament emission, milliamperes.

I_p = peak value of anode current impulse.

I_{av} = average anode current = D.C. component.

I_{gp} = peak value of grid current impulse.

$\mathcal{I}_a$ = amplitude of the fundamental component of anode current.

$\mathcal{V}_a$ = amplitude of oscillatory P.D. across load impedance.

E_a = anode supply voltage.

E_b = magnitude of grid bias voltage.

E_g = maximum permissible positive grid voltage.

E_o = minimum anode-filament P.D.

CHAPTER XI.—PARAS. 107-108

$\mathcal{V}_g$ = amplitude of grid excitation voltage.

R_d = dynamic resistance of load.

R_e = virtual load resistance = $\frac{R_d}{2}$.

μ = amplification factor of valve.

r_a = anode A.C. resistance of valve.

P_i = power input.

P_o = power output.

P_L = permissible dissipation of valve.

P_d = power actually dissipated.

P_g = grid driving power.

107. As already stated (Chapter IX) if I_p is the peak value of the semi-sinusoidal anode current impulse, its average value is $\frac{I_p}{\pi}$, and the amplitude of the fundamental (input frequency) oscillatory component will be $\frac{I_p}{2}$. The amplitudes of the respective harmonics are comparatively

small and are of little importance since the load impedance is a resonant circuit and is tuned to the input frequency. The fundamental component of the anode current is therefore the only one which can set up a considerable P.D. across the load impedance, and this P.D. is to all intents and purposes sinusoidal in spite of the impulsive nature of the anode current. As the oscillatory current and P.D. are proportional to the grid-filament voltage, the amplifier is said to be linear. The anode-filament P.D. is the difference between the supply voltage E_a and the instantaneous P.D. across the load. At the instant when the anode current reaches its peak value I_p , the anode-filament P.D. is $E_o = E_a - \mathcal{V}_a$. The oscillatory voltage acting in the anode circuit at this instant is $\mu \mathcal{V}_g - \mathcal{I}_a R_d$ and the anode current reaches the value

$$I_p = \frac{\mu \mathcal{V}_g - \mathcal{I}_a R_d}{r_a}$$

$$\text{Since } \mathcal{I}_a = \frac{I_p}{2},$$

$$I_p r_a = \mu \mathcal{V}_g - \frac{I_p R_d}{2}$$

$$= \mu \mathcal{V}_g - I_p R_e$$

$$I_p = \frac{\mu \mathcal{V}_g}{r_a + R_e}$$

The following special cases will now be considered :—

(i) The grid is not to be allowed to become positive with respect to the filament at any point in the cycle.

(ii) The grid is allowed to become positive with respect to the filament, but the effect of the resulting grid current upon the output and efficiency will be neglected.

Power relations without grid current

108. The conditions under (i) above are applicable when the power amplifier immediately follows a valve-driven master-oscillator (i.e. one without crystal or tuning fork control). If grid current is allowed to flow, the effective resistance of the "master" oscillatory circuit will vary during each cycle and will cause frequency variation. Even if the frequency is stabilized by

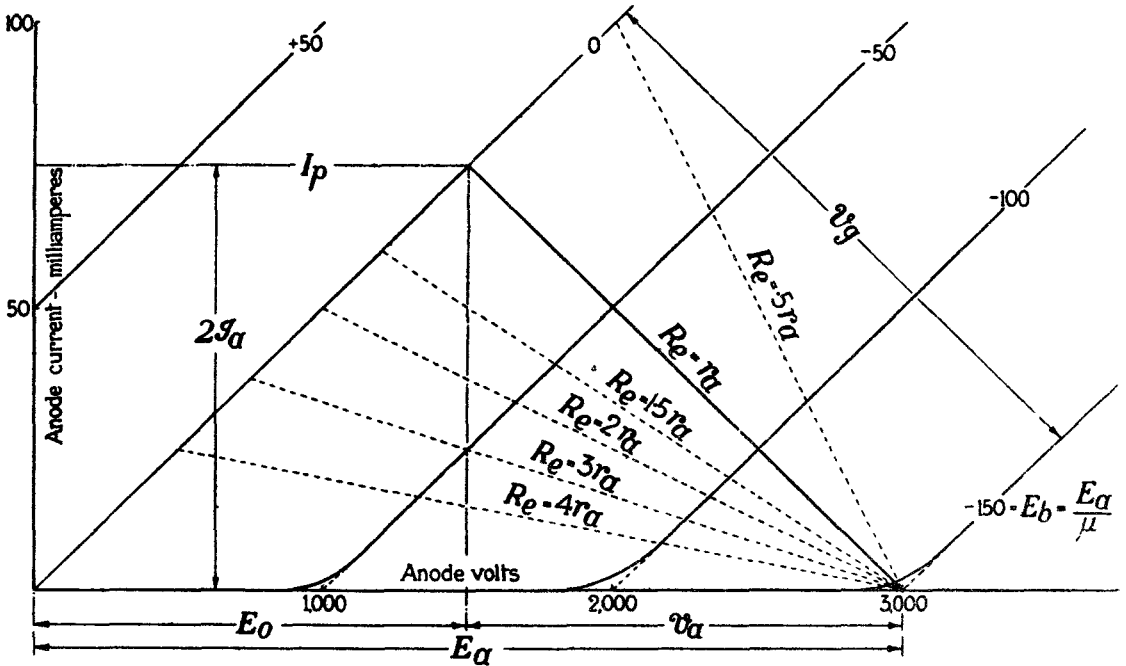


FIG. 49, CHAP. XI.—Class B amplifier; operation without grid current.

some form of mechanical oscillator this variation of load is to be avoided if possible. Referring now to fig. 49, the operating conditions are found as follows. The grid bias voltage is $-E_b = -\frac{E_a}{\mu}$, and the peak excitation will be numerically equal to the grid bias, i.e., $v_g = \frac{E_a}{\mu}$.

Then

$$I_p = \frac{E_a}{r_a + R_e}$$

$$g_a = \frac{E_a}{2r_a + R_d}$$

and the power delivered to the load is

$$P_o = \frac{g_a^2 R_d}{2} = \frac{E_a^2 R_d}{2(2r_a + R_d)^2}$$

$$= \frac{E_a^2 R_e}{4(r_a + R_e)^2}$$

This is a maximum when $R_e = r_a$ and the maximum possible power output is $\frac{E_a^2}{16r_a}$ watts. The power input P_i is equal to the product of the supply voltage and the average anode current.

CHAPTER XI.—PARA. 109

The latter is $\frac{I_p}{\pi}$ so that

$$P_i = \frac{E_a I_p}{\pi}$$

$$= \frac{E_a^2}{\pi (r_a + R_e)}$$

and the efficiency η is

$$\frac{P_o}{P_i} = \frac{R_e E_a^2}{4 (r_a + R_e)^2} \times \frac{\pi (r_a + R_e)}{E_a^2}$$

$$= \frac{\pi}{4} \frac{R_e}{r_a + R_e}$$

Thus the efficiency improves as the load impedance is increased, approaching the limiting value, 78·54 per cent., when R_e is very much larger than r_a . When, as in the present instance, the $I_a - V_a$ characteristics of the valve are available, the output and efficiency for any given load can be readily computed directly from the curves. In fig. 49 assuming that E_a (max.) = 3,000 volts, a load line representing a virtual resistance $R_e = r_a$ has been drawn in solid line, intersecting the curve $V_g = 0$ at $I_a = 75$, $V_a = 1,500$. Hence $I_p = 75$ milliamperes and $V_a = 1,500$ volts. The input, output and efficiency may now be calculated directly by means of the relations

$$P_o = \frac{V_a I_p}{4}$$

$$P_i = \frac{E_a I_p}{\pi}$$

$$\eta = \frac{P_o}{P_i} \times 100, \text{ per cent.}$$

Other load lines have also been drawn in dotted line, and the results derived from these are plotted in fig. 50. It is seen that the output rises to a maximum when $R_e = r_a$, but is very little reduced if $R_e = 2 r_a$, while the efficiency increases continuously towards its limiting value as R_e increases. For many purposes it is desirable to aim at a maximum value of the product ηP_o which gives the best compromise between output and efficiency. This quantity has also been plotted in fig. 50 and it is seen that a value of R_e in the region of from $1.5 r_a$ to $2.5 r_a$ satisfies this requirement.

Power when grid current is permitted

109. The output can be considerably increased by permitting grid current to flow, although some distortion of the anode current wave-form will be introduced. For the present, we shall continue to assume that the valve possesses ideal characteristics. Maximum output will obviously be obtained when both the anode current and the load P.D. have the greatest possible amount of variation. A further limitation must however be introduced, in that the anode-filament P.D. must never be allowed to fall below the instantaneous grid-filament P.D., i.e. the grid must not be allowed to become positive with respect to the anode. If this limitation is not observed, the secondary electrons emitted by the anode will be attracted to the grid and the grid current will be excessive. In order to allow for a reasonable amount of grid current, it may be taken that the peak anode current I_p must not exceed 80 per cent. of the filament emission. Even if the grid current reaches an instantaneous value of $\cdot 2 I_p$ the filament will then be able to supply the total current demanded by both electrodes. This total current is usually referred to as the space current. Referring now to fig. 51, the chain-dotted line has been drawn to indicate the positive limit of the grid swing—and therefore the limiting value of the minimum anode-filament P.D.—from the following considerations. Suppose the

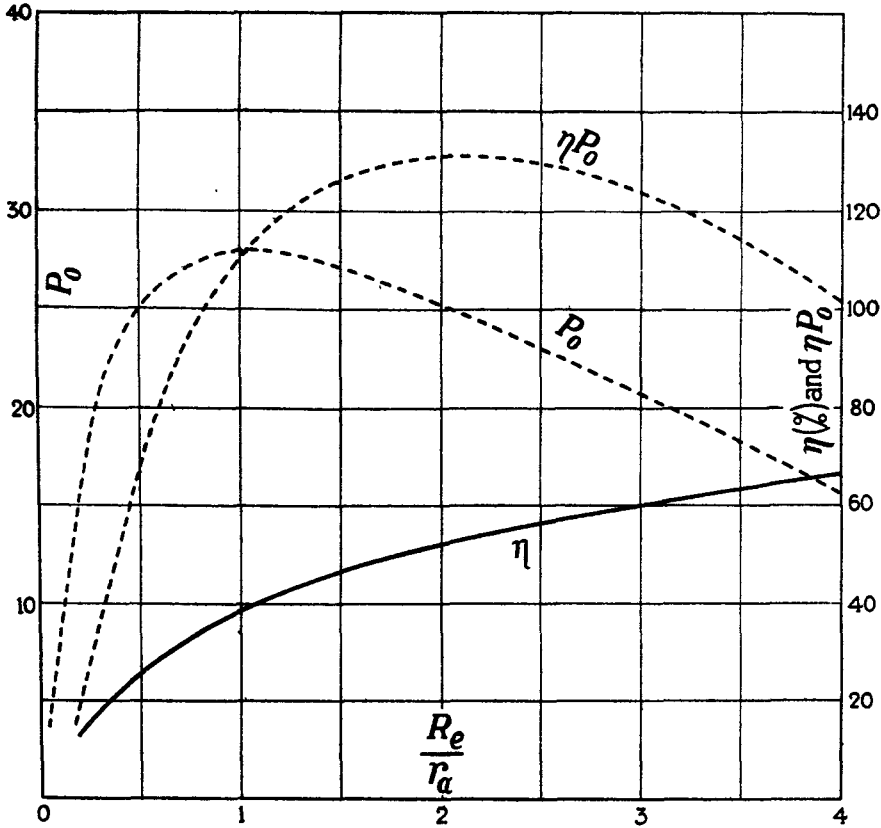


FIG. 50, CHAP. XI.—Output and efficiency of class B R.F. amplifier, for various ratios of $\frac{R_e}{r_a}$ and without allowing grid to become positive.

grid to be 200 volts positive with respect to the filament, then the anode must also be at least 200 volts positive, thus locating the point a. Similarly, when the grid is 100 volts positive, the anode filament P.D. must not be less than 100 volts, locating the point b. The chain-dotted line is drawn through these points. It is easily seen that for a given supply voltage E_a the greatest excursions of anode current and load P.D. will be achieved by locating the peak anode current I_p at its maximum permissible value, viz. $\cdot 8 I_e$ and operating on the load line corresponding to a virtual resistance

$$R_e = \frac{E_a - E_o}{I_p}$$

The value of E_o follows from the equation to the chain-dotted line. With the ideal characteristics postulated, the relation between I_a , V_a , and V_g is

$$I_a r_a = V_a + \mu V_g$$

When $V_a = E_o$, V_g is also equal to E_o and $I_a = I_p$, so that we may write

$$I_p r_a = E_o + \mu E_o$$

$$E_o = \frac{r_a}{\mu + 1} I_p$$

CHAPTER XI.—PARA. 109

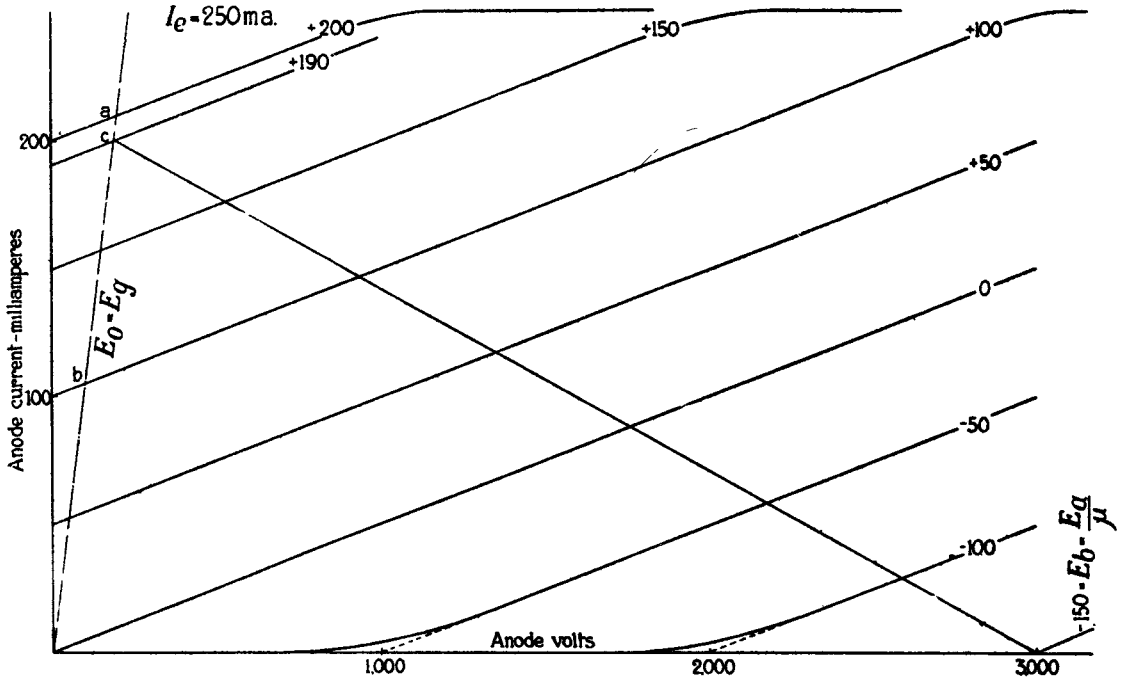


FIG. 51, CHAP. XI.—Class B amplifier; limiting E_0 line when grid current is permitted.

and the particular value of virtual load resistance corresponding to the load line is therefore

$$R_o = \frac{E_a}{I_p} - \frac{r_a}{\mu + 1}$$

In fig. 51, $I_p = 200$ milliamperes, $E_o = 190$ volts.

$$\begin{aligned} R_o &= \frac{3,000 - 190}{.2} \\ &= 14,050 \text{ ohms.} \end{aligned}$$

The output power is

$$\begin{aligned} P_o &= \frac{E_a - E_o}{4} I_p \\ &= \frac{E_a I_p}{4} - \frac{I_p^2 r_a}{4(\mu + 1)}, \end{aligned}$$

and in the particular example

$$\begin{aligned} P_o &= \frac{3,000 \times .2}{4} - \frac{.2 \times .2 \times 20,000}{4 \times 21} \\ &= 150 - 9.5 \\ &= 140.5 \text{ watts.} \end{aligned}$$

This may be verified directly from fig. 51, in which $E_a - E_o = 2,810$ volts, so that

$$P_o = \frac{2,810 \times .2}{4} = 140.5 \text{ watts as calculated.}$$

The input power is

$$\begin{aligned} P_i &= \frac{E_a I_p}{\pi} \\ &= \frac{3,000 \times .2}{\pi} \\ &= 191 \text{ watts,} \end{aligned}$$

and the efficiency 73.5 per cent. The actual power dissipated is only 50.5 watts.

The grid excitation is

$$\begin{aligned} \mathcal{V}_g &= \frac{E_a}{\mu} + E_g, \text{ and } E_g = E_o \\ \therefore \mathcal{V}_g &= \frac{E_a}{\mu} + \frac{I_p r_a}{\mu + 1}. \end{aligned}$$

In the example, $\frac{E_a}{\mu} = 150$, $\frac{I_p r_a}{\mu + 1} = 190$, and therefore $\mathcal{V}_g = 340$ volts. This again may be verified in Fig. 51, from which it is seen that the grid must reach + 190 volts from a bias of - 150 volts, so that the peak value of the oscillatory grid voltage must be $190 + 150 = 340$ volts as calculated. Finally, if the actual radio-frequency resistance R of the output circuit is known, the circulating current I_o in this circuit is found from the relation

$$\begin{aligned} P_o &= I_o^2 R \\ \text{or } I_o &= \sqrt{\frac{P_o}{R}}. \end{aligned}$$

Operation with limited power input

110. In certain circumstances it may be necessary to restrict the input power, e.g. to an amount not exceeding P_L , the permissible dissipation of the valve. Unless this condition is fulfilled a failure in the output circuit of the amplifier may result in damage to the valve. Provided that a sufficient grid swing is available, the greatest output and highest efficiency is obtained by operating with the maximum permissible anode supply voltage. The peak anode current is then given by the equation

$$I_p = \frac{\pi P_L}{E_a}$$

and the power output by

$$\begin{aligned} P_o &= \frac{(E_a - E_o) I_p}{4} \\ &= \frac{E_a I_p}{4} - \frac{I_p^2 r_a}{4(\mu + 1)} \\ &= \frac{\pi}{4} P_L - \frac{r_a}{4(\mu + 1)} \times \frac{\pi^2 P_L^2}{E_a^2}. \end{aligned}$$

The latter expression gives the power output in terms of the supply voltage and valve data and may therefore be used for calculation when the actual characteristics are not available.

CHAPTER XI.—PARA. 111

For example, taking the valve previously used for illustrative purposes, suppose that P_i is not to exceed P_L , i.e. 100 watts. If $E_a = 3,000$ volts as before,

$$\begin{aligned} I_p &= \frac{\pi \times 100}{3,000} \\ &= 104.6 \text{ milliamperes.} \\ P_o &= \frac{314}{4} - \frac{20,000}{84} \times \frac{\pi^2 \times 10^4}{9 \times 10^6} \\ &= 78.5 - 2.6 \\ &= 75.9 \text{ watts} \\ E_o &= \frac{I_p r_a}{\mu + 1} \\ &= 100 \text{ volts.} \end{aligned}$$

The virtual load resistance is

$$\begin{aligned} R_o &= \frac{r_a}{I_p} \\ &= \frac{3,000 - 100}{104.6 \times 10^{-3}} \\ &= 27,700 \text{ ohms.} \end{aligned}$$

To obtain this output, we must provide a grid excitation

$$\begin{aligned} V_g &= \frac{E_a}{\mu} + E_o \\ &= 150 + 100 \\ &= 250 \text{ volts.} \end{aligned}$$

Operation when limited by permissible dissipation

111. In some high-power ground station installations, arrangements are made for the valve or valves to be put out of action by an automatic device in the event of any failure which would cause a dangerous increase of anode current. The valves may then be operated in such a manner that they dissipate the power corresponding to their actual rating, although this is generally only possible if the valves and circuits are capable of withstanding a very high voltage. It is necessary to ensure that the anode is always at a potential considerably above that of the grid. Fig. 52 gives the $I_a - V_a$ characteristics of a 1,000-watt valve having an emission current of 1,200 milliamperes, a μ of 50 and an r_a of 20,000 ohms. The thin solid line O a represents the theoretical lower limit of the anode voltage swing, assuming that E_o must be not less than $5 E_g$. For ideal characteristics,

$$I_a r_a = V_a + \mu V_g$$

and when $I_a = I_p$, $V_a = E_o$ and $V_g = E_g = \frac{E_o}{5}$.

$$\therefore I_p = \frac{E_o}{r_a} \left(1 + \frac{\mu}{5} \right).$$

In general, if it is stipulated that $E_g = \frac{E_o}{n}$, $I_p = \frac{E_o}{r_a} \left(1 + \frac{\mu}{n} \right)$.

The actual characteristics are considerably curved in the region of low anode voltage, and the actual limiting line is that drawn in chain-dotted line. It is a close approximation to the theoretical one and the above equation may be used for practical calculation with negligible error. The greatest output will again be obtained with the highest permissible value of I_p . The dissipation is equal to the difference between the power input and output, i.e.

$$P_d = \frac{E_a I_p}{\pi} - \frac{V_a I_p}{4}$$

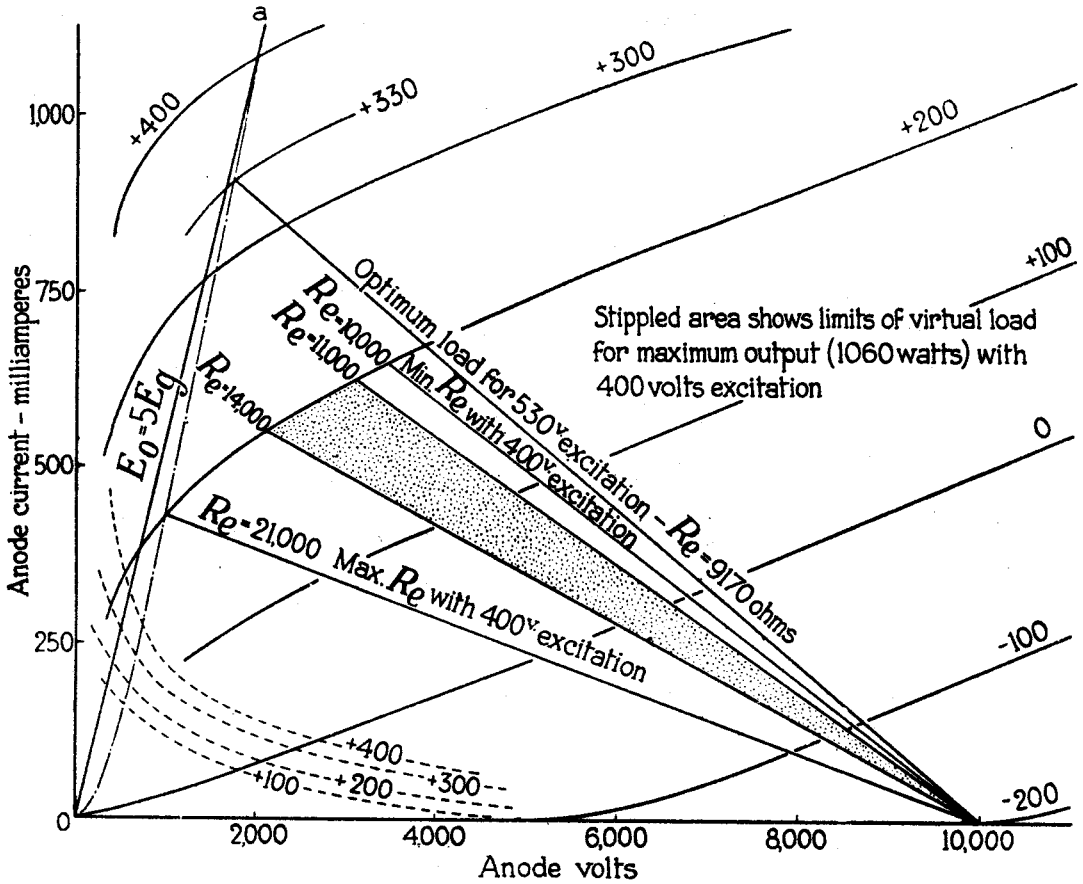


FIG. 52, CHAP. XI.—Operation of class B amplifier under different limiting conditions

If the valve is to dissipate its maximum power, $P_d = P_L$, while as it is also required to give maximum output, the virtual load will be approximately equal to r_a . Since $E_a = \mathcal{V}_a + E_o$,

$$\mathcal{V}_a I_p \left(\frac{1}{\pi} - \frac{1}{4} \right) + \frac{I_p^2 r_a}{\pi \left(1 + \frac{\mu}{n} \right)} = P_L$$

$$\frac{1}{\pi} - \frac{1}{4} = .0684,$$

$$.0684 \frac{\mathcal{V}_a}{I_p} + \frac{r_a}{\pi \left(1 + \frac{\mu}{n} \right)} = \frac{P_L}{I_p^2}$$

$$\frac{\mathcal{V}_a}{I_p} = R_e = r_a$$

$$I_p = \sqrt{\left\{ .0684 + \frac{1}{\pi \left(1 + \frac{\mu}{n} \right)} \right\} r_a \frac{P_L}{I_p^2}}$$

CHAPTER XI.—PARA. 112

With the valve under consideration, $\mu = 50$, $r_a = 20,000$, $P_L = 1,000$ and we have decided upon $n = 5$.

$$\begin{aligned}\therefore I_p &= \sqrt{\frac{1,000}{(\cdot 0684 + \cdot 028) 20,000}} \\ &= \cdot 716 \text{ ampere.}\end{aligned}$$

The operating conditions are therefore as follows:—

$$\begin{aligned}\mathcal{V}_a &= I_p R_o = \cdot 716 \times 20,000 \\ &= 14,320 \text{ volts.}\end{aligned}$$

$$\begin{aligned}E_o &= \frac{I_p r_a}{1 + \frac{\mu}{n}} = \frac{14,320}{11} \\ &= 1,300 \text{ volts.}\end{aligned}$$

$$E_a = 15,620 \text{ volts.}$$

If such a high supply voltage is actually employed, the insulation of the valve, and of certain portions of the circuit, will be called upon to withstand a potential difference of $E_a + \mathcal{V}_a = 29,940$ volts. If this is permissible, the power input will be

$$\begin{aligned}P_i &= \frac{E_a I_p}{\pi} = \frac{15,620 \times \cdot 716}{\pi} \\ &= 3,550 \text{ watts,}\end{aligned}$$

and the power output

$$\begin{aligned}P_o &= \frac{\mathcal{V}_a I_p}{4} = \frac{14,320 \times \cdot 716}{4} \\ &= 2,560 \text{ watts.}\end{aligned}$$

The actual dissipation is therefore

$$P_d = 990 \text{ watts}$$

and the efficiency

$$\eta = \frac{P_o}{P_i} = 72 \text{ per cent.}$$

The required grid excitation is

$$\begin{aligned}\mathcal{V}_g &= \frac{E_o}{n} + \frac{E_a}{\mu} \\ &= 260 + 332 \cdot 5 \\ &= 592 \cdot 5 \text{ volts.}\end{aligned}$$

112. Actually, the maximum permissible anode supply voltage for the valve having the characteristics of fig. 52 is only 10,000 volts. The permissible peak current may be derived from the relation previously used, viz.,

$$P_L = \frac{E_a I_p}{\pi} - \frac{\mathcal{V}_a I_p}{4}$$

or

$$E_a I_p \left(\frac{1}{\pi} - \frac{1}{4} \right) = P_L - \frac{I_p^2 r_a}{4 \left(1 + \frac{\mu}{n} \right)}$$

The permissible peak anode current is found by inserting the appropriate values of E_a , r_a , P_i , and n . With the valve of fig. 52, if $E_a = 10,000$ volts, we have

$$.0684 \times 10,000 I_p = 1,000 - 455 I_p^2$$

$$455 I_p^2 + 684 I_p - 1,000 = 0$$

$$I_p = 910 \text{ milliamperes.}$$

$$E_o = \frac{.91 \times 20,000}{11}$$

$$= 1,650 \text{ volts.}$$

$$\mathcal{V}_a = 10,000 - 1,650$$

$$= 8,350 \text{ volts}$$

$$R_o = \frac{\mathcal{V}_a}{I_p}$$

$$= \frac{8,350}{.91}$$

$$= 9,170 \text{ ohms.}$$

$$P_i = \frac{.91 \times 10,000}{\pi}$$

$$= 2,900 \text{ watts}$$

$$P_o = \frac{.91 \times 8,350}{4}$$

$$= 1,900 \text{ watts}$$

$$P_d = 2,900 - 1,900 = 1,000 \text{ watts} = P_L$$

$$\eta = \frac{1,900}{2,900}$$

$$= 65.5 \text{ per cent.}$$

$$\mathcal{V}_g = \frac{E_o}{n} + \frac{E_a}{\mu}$$

$$= 330 + 200$$

$$= 530 \text{ volts.}$$

Grid driving power

113. The curves shown in the bottom left-hand corner of fig. 52 are the $I_g - V_a$ characteristics for different values of V_g . With their aid it is possible to form an estimate of the power which must be expended in driving the grid positive as in the previous example. It will be seen that at the instant when the anode current reaches the value I_p the grid current also reaches its peak value I_{gp} , about 158 milliamperes. The average grid current may be taken as $\frac{I_{gp}}{2}$, and the grid driving power is approximately equal to one-half the product of the average grid current and the peak excitation, i.e.

$$P_g = \frac{I_{gp} \mathcal{V}_g}{4}$$

CHAPTER XI.—PARA. 114

Under the operating conditions of the previous paragraph,

$$P_g = \frac{.158 \times 530}{4}$$

$$= 21 \text{ watts.}$$

It will now be seen that one reason for placing a fairly high limiting value upon E_o , i.e. $E_o = 4 E_g$ or $5 E_g$ rather than $E_o = E_g$, is to reduce the power required for the grid drive. Thus, suppose that the valve of fig. 52 is operated under the latter conditions, with such a supply voltage and anode load that $I_p = 875$ milliamperes when $E_g = E_o = 400$ volts. The grid will at the same instant call for 400 milliamperes, so that the total space current should be 1,275 milliamperes.

The filament is unable to supply this, but will give only $\frac{875}{1,275} \times 1,200 = 824$ milliamperes to the anode circuit and $\frac{400}{1,275} \times 1,200 = 376$ milliamperes to the grid. The power output will therefore be less than that calculated without taking the grid current into account, while the grid driving power will have to be considerably increased, i.e. if the grid excitation is 530 volts as before, to 50 watts.

Output with limited excitation

114. If both the supply voltage and the available excitation are limited, the load resistance for maximum output may be considerably less than r_a , and its value somewhat critical. This is due to the curvature of the characteristics and is most easily appreciated by laying down a set

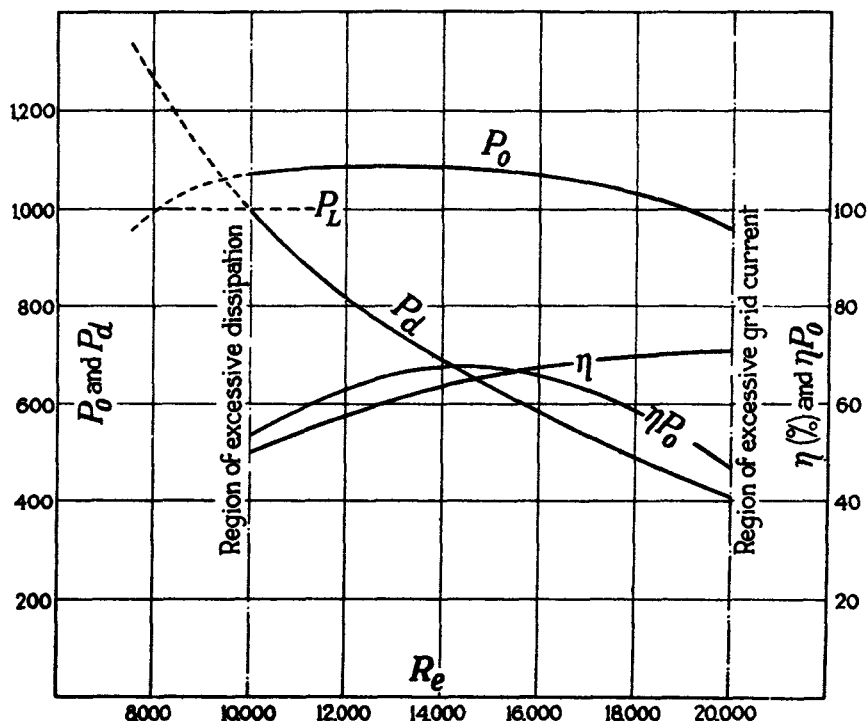


FIG. 53, CHAP. XI.—Output and efficiency of class B R.F. amplifier, with limited grid excitation, for various values of virtual load resistance.

of conditions and calculating the output with different values of R_e . Taking the valve characteristics of fig. 52 let $E_a = 10,000$ volts, $E_g \gg \frac{E_o}{5}$ and suppose that V_g cannot exceed 400 volts. Then one end of the virtual load line will be located at $I_a = 0$, $V_a = E_a = 10,000$, and the other will lie somewhere on the curve $V_g = +200$. Owing to the curvature in the region of low anode voltage, it is impossible to operate with a virtual load greater than r_a (approximately) unless either the excitation is reduced or the grid allowed to become more positive than $\frac{E_o}{5}$. On the other hand if a comparatively low virtual load is adopted, i.e. less than $\frac{r_a}{2}$, the dissipation becomes excessive and the output again falls off. The optimum virtual

load is approximately equal to the anode A.C. resistance of the valve, measured at the point where the load line reaches the $I_a - V_a$ curve corresponding to the positive limit of grid voltage, but this point can only be located by trial and error. The manner in which the power output varies with the load resistance, for a peak excitation of 400 volts and an anode voltage of 10,000 volts, is shown in fig. 53. It is seen that if R_e falls below 10,000 ohms the permissible dissipation will be exceeded, that values of R_e between 10,000 and 16,000 ohms give substantially the same output, and that the product ηP_o is a maximum when R_e is about 15,000 ohms. On the whole, it appears desirable to operate under the latter conditions, since the actual power dissipated in the valve is then much below the permissible value.

Anode current wave-form

115. It has hitherto been assumed that in a class B amplifier the wave-form of the anode current impulse is truly a half sine wave, but this is seldom if ever so. The degree of departure depends upon the amount by which the actual valve characteristics differ from those of an ideal valve having the same constants. The wave-form of the impulse set up by a sinusoidal excitation voltage can be derived from the $I_a - V_a$ characteristics in the following manner. Referring to fig. 54, in which the curves from $V_g = +5$ to $V_g = -3$ are given, the appropriate virtual load line is first drawn, and this is then divided into a number of parts equal to the number of curves, i.e. eight. From the point 0, the curves are lettered A B C . . . H, and the equal divisions on the load line a b c . . . h. If all eight curves from $V_g = -2$ to $V_g = +5$ were equally spaced, a sinusoidal excitation voltage would set up a semi-sinusoidal anode current

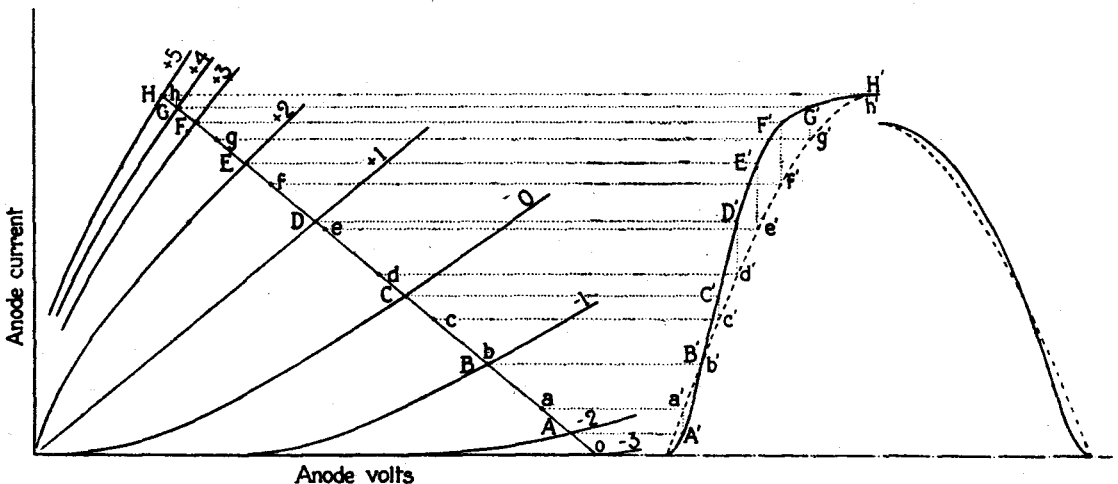


FIG. 54, CHAP. XI.—Distortion due to curvature of characteristics.

CHAPTER XI.—PARA. 116

impulse, as shown immediately to the right of the characteristics, the points $a' b' c' \dots h$, being the projections of the points $a b c \dots h$, as indicated by the dotted lines. As it is, however, the grid is initially biased to -3 volts, and on reaching -2 volts, instead of rising to the value corresponding to a , the current rises only to A . The point A' which is obtained by projecting A until it lies vertically below a' is a point on the current wave-form. When the grid voltage rises to -1 volt, the points B and b coincide, as also do their projections B' and b' . Similarly the projection C' of the point C lies vertically above the projection C' of the point c . The construction of the current wave-form $A' B' C' \dots H'$ is obtained by repetition of the process described. It is seen that the wave is rather more flat-topped than a true sine curve. In practice a class B amplifier is rarely operated under conditions giving such a high degree of distortion as this; the characteristics corresponding to $V_g = +3$, $V_g = +4$ and $V_g = +5$ are deliberately drawn close together in order to exaggerate the effect for illustrative purposes. The grid is also shown as being biased slightly beyond the cut-off voltage, giving rise to distortion of the lower portion of the current wave. If the grid swing were limited in such a manner that the grid reached a maximum positive potential of only 3 volts, the current wave-form would be that shown in the extreme right of the diagram, with the corresponding half sine wave in dotted line for comparison. This is more closely representative of practical operating conditions.

Amplification of modulated waves

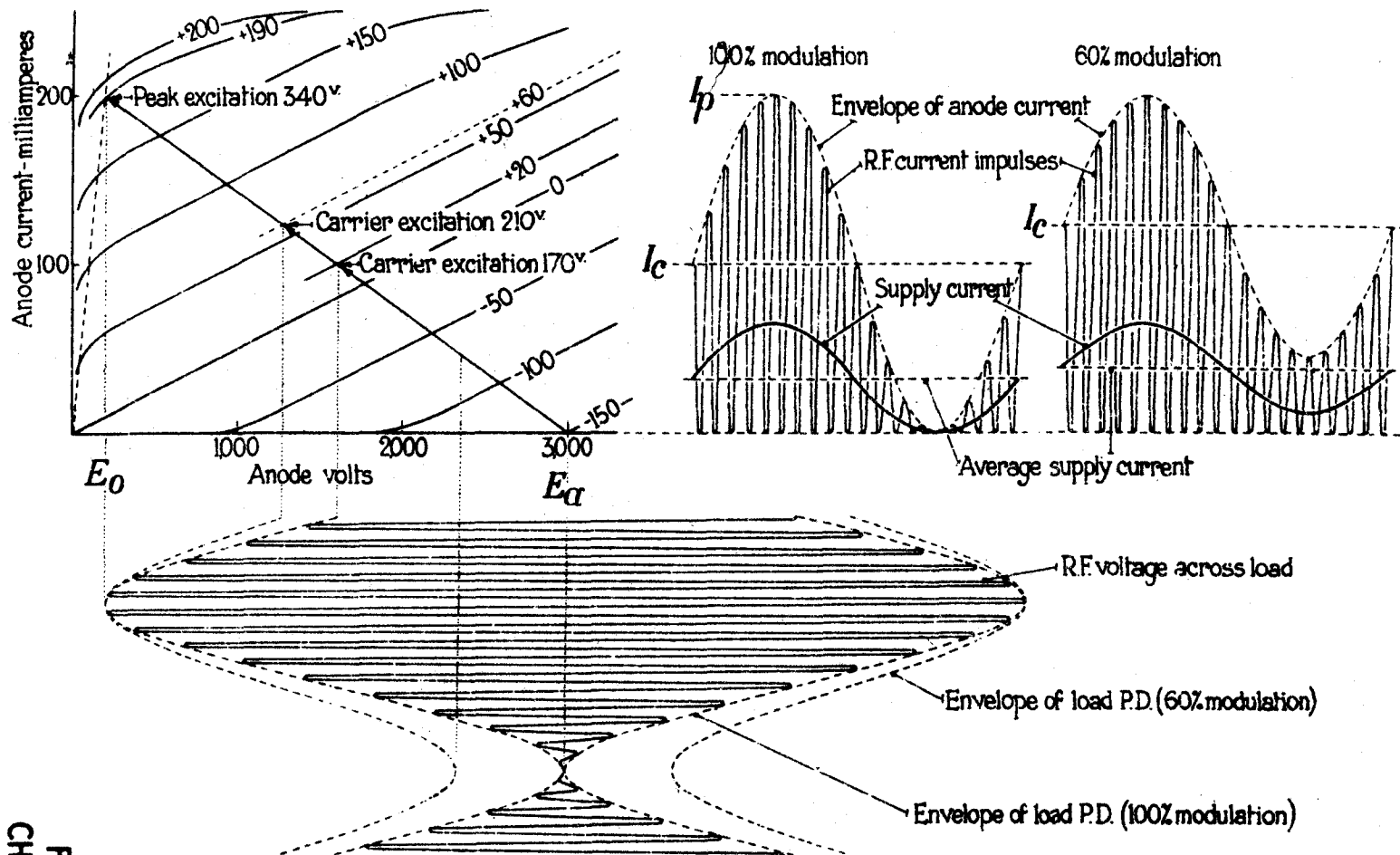
116. The principal use of the class B radio-frequency power amplifier is for the amplification of modulated waves. Although the wave-form of the anode current impulse is not identical with that of the grid excitation voltage, the envelopes of the anode current and load P.D. will be nearly so provided that the operating conditions are chosen in such a manner that the dynamic characteristic is sensibly straight. This implies that the dynamic resistance of the load should be as large as the other conditions will allow, and as already shown, R_d should if possible be not less than $2r_a$ and preferably of the order of $4r_a$. It must, however, be observed that a high dynamic load resistance can be profitably employed only if a very high supply voltage is available, and the grid excitation must also be correspondingly increased. If the relation between anode current and excitation voltage were truly linear, the average anode current would remain constant, irrespective of the depth of modulation, and the reading of a milliammeter in the anode circuit is therefore an indication of the degree with which linearity is approached. In discussing modulated waves we are chiefly interested in (i) the carrier conditions and (ii) the conditions at the peak of the modulation cycle. It is convenient to denote quantities such as peak anode current, load P.D., etc., under the former conditions by I_p , $\mathcal{V}_a$, etc., and the corresponding quantities at the modulation peak by I'_p , $\mathcal{V}'_a$, etc. Assuming that provision is to be made for 100 per cent. sinusoidal modulation, the peak anode current I'_p when completely modulated will be found as above, but the peak anode current I_p of the unmodulated carrier will be equal to $\frac{I'_p}{2}$. The load P.D. will also be only one-half its greatest permissible value, hence

the input power P_i with unmodulated carrier, will be $\frac{E_a I_p}{\pi} = \frac{E_a I'_p}{2\pi}$, and the carrier output

power P_o will be $\frac{(E_a - E_o) I_p}{4} = \frac{(E_a - E_o) I'_p}{8}$. For example, referring to fig. 55, with an

anode supply voltage of 3,000 volts and a virtual load of 14,050 ohms, the peak value of the anode current under unmodulated conditions will be 100 milliamperes, and the input power $\frac{E_a I_p}{\pi} = \frac{3,000 \times .1}{\pi} = 95.5$ watts. The carrier power output will be $\frac{2,810 \times .1}{8} = 35.125$ watts,

and the efficiency $\frac{35.125}{95.5} = 36.7$ per cent. Comparing these results with those of para. 109, it will be observed that the power output is one-quarter and the efficiency one-half of that obtained



OPERATION OF LINEAR AMPLIFIER

with the same supply voltage and load impedance, but operating with the greatest permissible grid swing, i.e. 340 volts. The grid excitation required to produce this carrier power will of course be only 170 volts.

117. If the wave is to be modulated to a depth of less than unity, the carrier power may be increased. Thus if the load resistance and supply voltage remain as before, viz. $R_e = 14,050$ ohms, $E_a = 3,000$ volts, and the excitation under carrier conditions is increased to 210 volts, $I_p = 125$ milliamperes. The load P.D. will now be 1,760 volts and the power output $\frac{1,760 \times .125}{4} = 55$ watts, while the power input is $\frac{3,000 \times .125}{\pi} = 119$ watts. The carrier efficiency is now 46 per cent. Under these conditions the greatest anode current change which can be achieved, without departure from linearity, is 75 milliamperes. The corresponding depth of modulation is easily found ;

$$I_p = 125 + 75 = 200 \text{ milliamperes}$$

$$\mathcal{I}'_a = \frac{I'_p}{2} = 100 \text{ milliamperes}$$

$$I_p = 125 \text{ milliamperes}$$

$$\mathcal{I}_a = 62.5 \text{ milliamperes}$$

$$K = \frac{\mathcal{I}'_a - \mathcal{I}_a}{\mathcal{I}_a}$$

$$= \frac{100 - 62.5}{62.5}$$

$$= .6,$$

corresponding with a depth of 60 per cent.

118. Since, if the amplifier is truly linear, the average current during modulation is the same as when modulation is not occurring, the power input is the same no matter what the depth of modulation may be, whereas the output power during modulation is greater than under unmodulated conditions. Thus when modulation is taking place, the power converted into heat is greater than during unmodulated periods but the power output is greatly increased and the efficiency is increased. For this reason, the dynamic resistance of the load should always be adjusted by an alteration of anode tapping point (or rearrangement of the tappings on the pulse coil, where fitted) with an unmodulated grid excitation corresponding to the desired carrier amplitude. In order to obtain a high value for the product ηP_o the virtual resistance of the load should be at least equal to r_a and preferably greater than $1.5 r_a$. Care must be taken however, not to exceed the maximum permissible supply voltage, which must increase with the load resistance in order to maintain the desired output.

The class C amplifier

119. The class C amplifier is employed for the following purposes :—

(i) In all the amplifier stages of W/T transmitters. In the stage immediately following the master-oscillator it is preferable to avoid running into grid current, in order to reduce the load on the master-oscillator and to maintain it at a constant magnitude during the whole oscillatory cycle.

(ii) As the modulated stage in a frequency-controlled R/T transmitter. The amplifier may be modulated by imposing an audio-frequency variation either upon the anode voltage or upon the grid-filament P.D., the former being referred to as an anode-modulated amplifier, and the latter as a grid-modulated amplifier. The physical aspects of the process of modulation are dealt with in Chapter XII.

(iii) As a frequency multiplier, the amplified output being usually of double or treble the input frequency. The stage is then called a frequency-doubler or frequency-trebler.

Class C W/T amplifier

120. As already stated, in the class C amplifier the grid is biased to beyond cut-off point and the anode current flows for less than 180 electrical degrees; the actual angle is generally between 120° and 160° . It is convenient to use one-half of this angle for purposes of calculations and in the following discussion, the symbol φ will be used to denote one-half the angle of anode current flow; φ will be referred to as the "operating angle". For example, in a class B amplifier $\varphi = 90^\circ$. It is easily seen that the quantities $\mathcal{V}_g$, E_g , E_b , E_a and φ are interdependent. In all practical cases the maximum anode supply voltage is fixed by practical considerations, while the minimum anode-filament P.D. E_o and most positive grid potential E_g must be decided as in any other form of amplifier. Thus it is usually necessary to decide upon a tentative value of φ and find the relations between $\mathcal{V}_g$, E_b and φ .

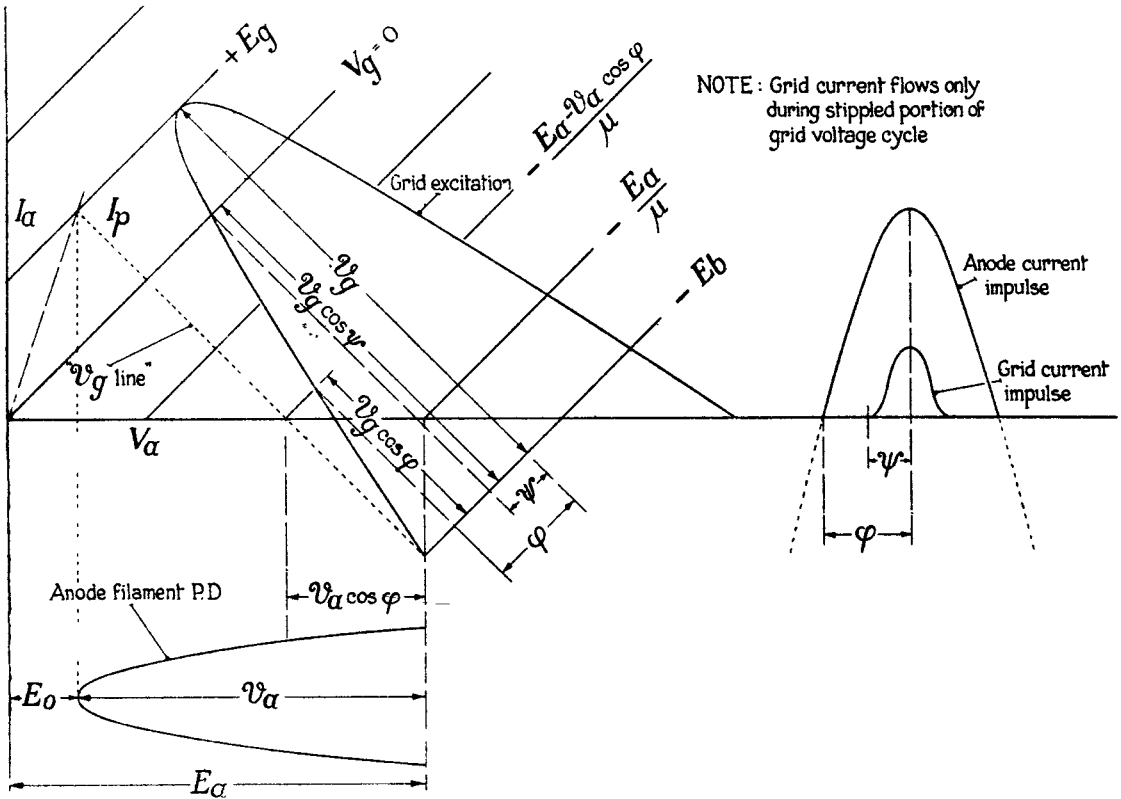


FIG. 56, CHAP. XI.—Relation between various operating voltages in class C amplifier.

Referring to fig. 56, the anode supply voltage being E_a and the minimum anode-filament P.D. E_o , the load P.D. will be $\mathcal{V}_a = E_a - E_o$. It is seen that anode current will not commence to flow until the instantaneous grid voltage has arisen from $-E_b$ to $-E_b + \mathcal{V}_g \cos \varphi$. The oscillatory load P.D. v_a rises as the grid becomes less negative and at the instant when anode current commences, $v_a = \mathcal{V}_a \cos \varphi$. At this instant the grid potential is $\frac{E_a - \mathcal{V}_a \cos \varphi}{\mu}$. Assuming that the grid may be allowed to swing positive up to the value E_g volts therefore, the required excitation is

$$\mathcal{V}_g = E_g + \frac{E_a - \mathcal{V}_a \cos \varphi}{\mu} + \mathcal{V}_g \cos \varphi$$

or

$$\mathcal{V}_g = \left[E_g + \frac{E_a - \mathcal{V}_a \cos \varphi}{\mu} \right] \frac{1}{1 - \cos \varphi}.$$

The magnitude of the required bias voltage is

$$\begin{aligned}
 E_b &= \mathcal{V}_g - E_g \\
 &= \left[E_g + \frac{E_a - \mathcal{V}_a \cos \varphi}{\mu} \right] \frac{1}{1 - \cos \varphi} - E_g \\
 &= E_g \frac{\cos \varphi}{1 - \cos \varphi} + \frac{E_a}{\mu} \left(\frac{1}{1 - \cos \varphi} \right) - \frac{(E_a - E_0) \cos \varphi}{\mu (1 - \cos \varphi)} \\
 &= \frac{E_a}{\mu} + \left(E_g + \frac{E_0}{\mu} \right) \frac{\cos \varphi}{1 - \cos \varphi}.
 \end{aligned}$$

This relation is of fundamental importance in the theory of class C amplification. The sign of E_b is of course negative.

Power output and efficiency

121. Since the anode current flows for less than 180° , the anode current impulses are no longer half sine waves, so that the fundamental component (H_1) of the anode current will be less than $\frac{I_p}{2}$ and the average anode current I_{av} less than $\frac{I_a}{\pi}$. It is possible to calculate the ratios

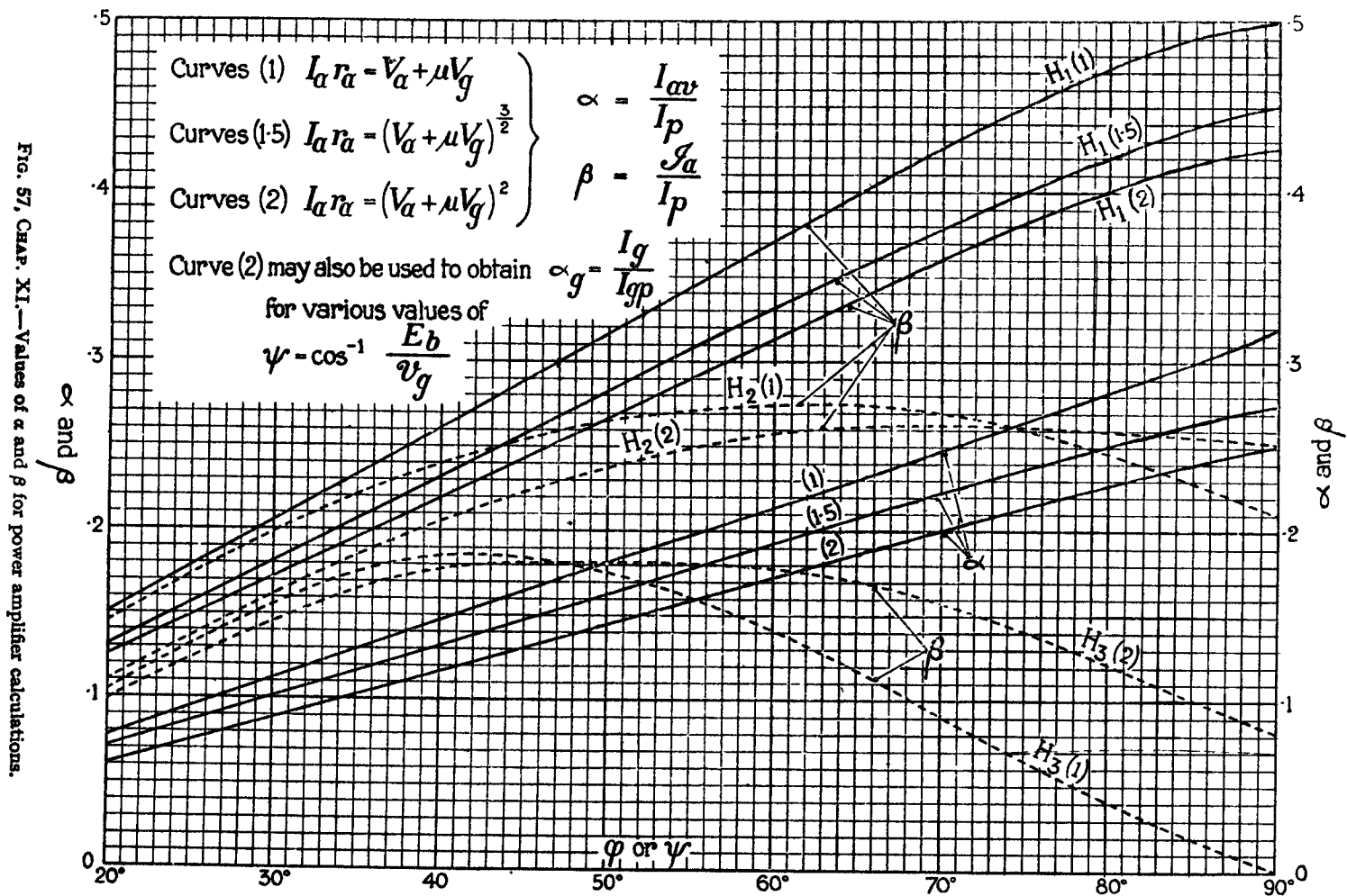
$\frac{I_{av}}{I_p} = \alpha$ and $\frac{\mathcal{I}_a}{I_p} = \beta$ for various operating angles. The results are exhibited in graphical form in fig. 57, in which three pairs of H_1 curves are given in solid line. Those marked (1) are calculated for ideal characteristics, i.e. assuming that $r_a I_a = V_a + \mu V_g$. Similarly the curves marked (1.5) refer to characteristics obeying the law $r_a I_a = (V_a + \mu V_g)^{1.5}$ and those marked (2) to characteristics obeying the law $r_a I_a = (V_a + \mu V_g)^2$. The latter are also of use in determining the grid driving power with greater accuracy than has hitherto been attempted. With the aid of these curves the power input, output and efficiency can be determined by means of the relations

$$\begin{aligned}
 P_i &= I_{av} E_a \\
 &= \alpha I_p E_a \\
 P_o &= \frac{\mathcal{I}_a \mathcal{V}_a}{2} \\
 &= \frac{\beta I_p \mathcal{V}_a}{2} \\
 \eta &= \frac{P_o}{P_i}.
 \end{aligned}$$

122. To illustrate the method of calculating the performance of a class C amplifier we may take the 1,000-watt valve previously discussed. Allowing for grid current we have already found that with the limiting value of E_0 set at not less than $5 E_g$ the maximum permissible peak anode current is about 900 milliamperes, while the maker stipulates that E_a shall not exceed 10,000 volts. Assuming that an operating angle of 75° is chosen and that the first power law is obeyed, $\alpha = .265$, $\beta = .45$. The maximum permissible input is therefore

$$\begin{aligned}
 P_i &= \alpha I_p E_a = .265 \times .9 \times 10,000 \\
 &= 2,380 \text{ watts,}
 \end{aligned}$$

and the minimum value of E_0 is found directly from the valve characteristics to be 1,800 volts.

Fig. 57, Chap. XI.—Values of α and β for power amplifier calculations.

Hence $\mathcal{V}_a$ will be 8,200 volts.

$$\begin{aligned} P_o &= \frac{\mathcal{I}_a \mathcal{V}_a}{2} = \frac{\beta I_p \mathcal{V}_a}{2} \\ &= .225 \times .9 \times 8,200 \\ &= 2,050 \text{ watts.} \\ P_d &= P_i - P_o \\ &= 330 \text{ watts.} \\ \eta &= 86 \text{ per cent.} \end{aligned}$$

The appropriate load impedance R_d is equal to

$$\frac{\mathcal{V}_a}{\mathcal{I}_a} = \frac{\mathcal{V}_a}{\beta I_p} = 40,700 \text{ ohms.}$$

Since $\cos 75^\circ \doteq .26$ and $E_g = \frac{E_o}{5}$, the required grid bias will be

$$\begin{aligned} E_b &= \frac{10,000}{50} + \left(\frac{1,800}{5} + \frac{1,800}{50} \right) \frac{.26}{1 - .26} \\ &= 200 + 139 \\ &= 339 \text{ volts.} \end{aligned}$$

The grid excitation necessary to obtain this output is

$$\begin{aligned} \mathcal{V}_g &= \left(\frac{1,800}{5} + \frac{10,000 - 8,200 \times .26}{50} \right) \frac{1}{1 - .26} \\ &= 705 \text{ volts.} \end{aligned}$$

It is of interest to repeat the calculations on the assumption that the characteristics obey the $\frac{3}{2}$ -power law. From curves (1.5) of fig. 57, if $\varphi = 75^\circ$, $\alpha = .234$ and $\beta = .4$. Hence for the same value of I_p and supply voltage E_a ,

$$\begin{aligned} P_i &= .234 \times .9 \times 10,000 \\ &= 2,106 \text{ watts,} \\ P_o &= .2 \times .9 \times 8,200 \\ &= 1,476 \text{ watts,} \\ P_d &= 630 \text{ watts,} \\ \eta &= 70 \text{ per cent.} \end{aligned}$$

The load impedance under these conditions will be

$$\frac{\mathcal{V}_a}{\beta I_p} = \frac{8,200}{.4 \times .9} = 22,800 \text{ ohms.}$$

Grid driving power

123. Referring again to fig. 56 it is seen that grid current commences when $\mathcal{V}_g \cos \psi = E_b$ and therefore flows during an interval corresponding to an operating angle ψ , where

$$\cos \psi = \frac{E_b}{\mathcal{V}_g}$$

CHAPTER XI.—PARAS. 124–126

In the previous example therefore

$$\cos \psi = \frac{339}{705} = .48,$$
$$\psi = 61^\circ.$$

Note that this angle is smaller than ϕ . The grid current impulse is in fact considerably more peaky than a sine curve; the peak grid current I_{gp} can be found from the $I_g - V_a$ characteristics and its average value I_g is approximately given by

$$I_g = a_g I_{gp}$$

where a_g is taken from the square law curve (2) of fig. 57, using the angle ψ instead of ϕ . In the present example $I_m = 158$ milliamperes and $a_g = .175$, the grid driving power being

$$P_g = a_g I_{gp}^2 \mathcal{V}_g$$
$$= .175 \times .158 \times 705$$
$$= 19.5 \text{ watts.}$$

The anode-modulated amplifier

124. The operation of this amplifier resembles that of the W/T amplifier in that the grid excitation is derived from a master-oscillator (or a buffer amplifier) and is of constant amplitude during the whole of the audio-frequency cycle. The anode voltage is the sum of the supply voltage E_a and the audio-frequency voltage set up in the speech choke, the amplitude of the latter voltage being dependent upon the intensity of the sound input. The greatest audio-frequency voltage which can be handled without serious distortion possesses an amplitude equal to the anode supply voltage, thus the total H.T. voltage varies between zero and $2E_a$ during the audio-frequency cycle. In order to obtain distortionless modulation the relation between currents and voltages must be linear, i.e. when the supply voltage is doubled, the P.D. across the load impedance, and the fundamental component of anode current, must also rise to twice the values under unmodulated conditions. Averaged over an audio-frequency cycle however, the supply voltage and current remain nearly constant because the superimposed variations are practically symmetrical about their mean values. For sinusoidal modulation to a depth of unity the power input must be 1.5 times the carrier input, and the output power must also increase by 50 per cent., in order to supply the side-band power. The average efficiency does not vary to any extent and therefore the power dissipated during sinusoidal modulation will be about 50 per cent. greater than under carrier conditions.

125. In general the anode-modulated amplifier requires a much higher grid bias than the W/T amplifier, because the grid bias must be allowed to vary during the modulation cycle. It is desirable that this variation should not be too large a fraction of the total bias, and therefore the latter should be as large as possible. It follows that the operating angle should be considerably smaller than in the W/T amplifier.

126. An idea of the possibilities with regard to any particular valve is most easily obtained by a study of the operation, first under carrier conditions, and afterwards at the peak of the audio-frequency cycle. The process will be illustrated with reference to the valve of which the ideal characteristics are given in fig. 58. The valve constants are, $r_a = 10,000$ ohms, $\mu = 20$, $I_e = 800$ milliamperes, $P_L = 60$ watts. The carrier conditions will be first investigated, the notation used being as in previous paragraphs. In a low-power valve of this kind we may allow the anode-filament P.D. to fall to not less than the maximum positive grid voltage, and the chain-dotted line represents its lower limit, $E_o = E_g$. As a W/T amplifier the maximum permissible supply voltage is 3,000 volts, but as an anode-modulated amplifier it is assumed

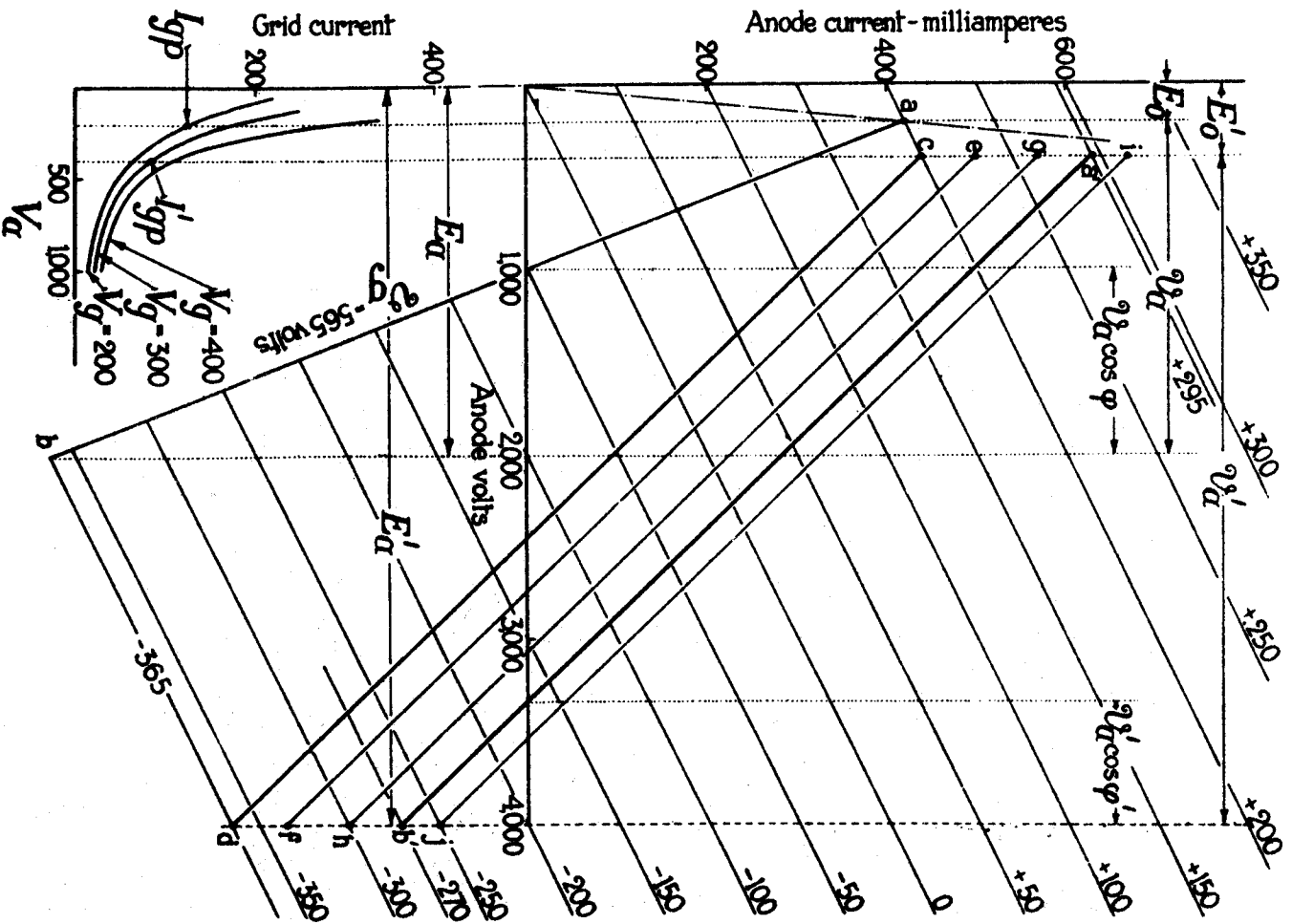


FIG.58
OPERATION OF ANODE MODULATED AMPLIFIER
CHAP. XI

to be only 2,000 volts. If then we also stipulate that E_o shall not be less than 10 per cent. of the supply voltage, since $E_a = 2,000$ volts, $E_o = 200$ volts. The value of I_p follows from the equation to the chain-dotted line, i.e.

$$\begin{aligned} I_p &= E_o \frac{\mu + 1}{r_a} \\ &= \frac{200 \times 21}{10,000} \\ &= 420 \text{ milliamperes.} \end{aligned}$$

It is now necessary to adopt a trial operating angle which will give a sufficiently large output with a high efficiency and a dissipation not exceeding $\frac{2}{3} P_L$. A value of ϕ in the neighbourhood of 60° is indicated in order that the carrier bias may be considerably negative, i.e. of the order of three times the "cut-off" bias. The latter is equal to $-\frac{E_a}{\mu}$ or -100 volts. Actually the line a b, representing the excursion of the excitation voltage, was drawn from a through the point $V_a = 1,000$, $I_a = 0$, intersecting the E_a line at $V_g = -365$. Then

$$\begin{aligned} \mathcal{V}_a \cos \phi &= 1,000 \\ \mathcal{V}_a &= 1,800 \\ \cos \phi &= \frac{1,000}{1,800} \\ &= .554, \end{aligned}$$

so that $\phi = 56^\circ$ (nearly).

A trial calculation of the input, output and efficiency then shows whether this bias and operating angle is satisfactory. From fig. 57 for $\phi = 56^\circ$, $\alpha = .2$, $\beta = .348$

$$\begin{aligned} I_{av} &= \alpha I_p \\ &= .2 \times 420 \\ &= 84 \text{ milliamperes.} \\ \mathcal{I}_a &= \beta I_p \\ &= .348 \times 420 \\ &= 146 \text{ milliamperes.} \\ P_i &= I_{av} E_a \\ &= .084 \times 2,000 \\ &= 168 \text{ watts.} \\ P_o &= \frac{\mathcal{V}_a \mathcal{I}_a}{2} \\ &= \frac{146 \times 1,800}{2} \\ &= 131.4 \text{ watts.} \\ P_d &= P_i - P_o \\ &= 36.6 \text{ watts,} \end{aligned}$$

CHAPTER XI.—PARA. 127

which is rather less than $\frac{2}{3} P_L$ and may be considered satisfactory. The efficiency is

$$\begin{aligned}\eta &= \frac{P_o}{P_i} \\ &= \frac{131.4}{168} \\ &= 78 \text{ per cent.},\end{aligned}$$

and the grid excitation required for these conditions is

$$\mathcal{V}_g = E_b + E_g = 365 + 200 = 565 \text{ volts.}$$

127. In order to obtain 100 per cent. distortionless modulation the audio-frequency voltage must, during its cycle, vary the anode voltage of the modulator valve between 0 and $2E_a$. At the peak of the cycle, therefore, $E'_a = 2E_a$; it is also necessary that $\mathcal{V}'_a = 2\mathcal{V}_a$ and $\mathcal{J}'_a = 2\mathcal{J}_a$; in other words the instantaneous power output at the peak must be four times the carrier output. It follows that $E'_o = 2E_o$. The grid excitation, being derived from the previous stage, will remain constant at 565 volts. The slope of the line representing the excitation—which may be called the “ $\mathcal{V}_g$ -line”—is easily found, since one end must lie on the vertical through $E'_a = 4,000$, and it must extend over a range of 565 volts. Several trial $\mathcal{V}_g$ -lines are drawn in the diagram. The first, (C d) was inserted in order to deduce what would occur if the grid bias were held rigidly constant. It crosses the anode voltage base line at $V_a = 2,450$ volts, so

that $\mathcal{V}'_a \cos \varphi' = 1,650$ volts. It follows that $\cos \varphi' = \frac{1,650}{3,600} = .457$, $\varphi' = 63^\circ$. Referring to

fig. 57, for $\varphi' = 63^\circ$, $\beta = .388$. On this $\mathcal{V}_g$ -line the value of I'_p is 440 milliamperes, therefore $\mathcal{J}'_a = .388 \times 440 = 172$ milliamperes, which is much less than $2\mathcal{J}_a$. Thus an operating angle of 63° , at the audio-frequency peak, will not give distortionless modulation. Successive trials with the $\mathcal{V}_g$ -lines e f and g h gave the following operating angles and values of $\mathcal{J}'_a$:—

line e f, $I'_p = 500$, $\varphi' = 70^\circ$, $\mathcal{J}'_a = 212$ milliamperes ;

line g h, $I'_p = 570$, $\varphi' = 74^\circ$, $\mathcal{J}'_a = 254$ milliamperes ;

line i j, $I'_p = 670$, $\varphi' = 82^\circ$, $\mathcal{J}'_a = 322$ milliamperes.

The latter is fairly close to the desired value and, after a further trial, too close to the final value to be shown, the line a' b' was chosen. Here $I'_p = 630$, $E'_b = 270$, $\mathcal{V}'_a \cos \varphi' = 650$, $\cos \varphi' = .18$ and $\varphi' = 79^\circ$. From fig. 57, $\beta' = .468$ and therefore $\mathcal{J}'_a = .468 \times 630 = 295$ milliamperes. This is quite near to the desired value (292 milliamperes) and the remainder of the calculations may be completed. For $\varphi' = 79^\circ$, $\alpha' = .27$, and therefore

$$\begin{aligned}I'_{av} &= \alpha' I'_p = 27 \times 630 \\ &= 170 \text{ milliamperes}\end{aligned}$$

$$\begin{aligned}P_i &= I'_{av} E'_a \\ &= .170 \times 4,000 \\ &= 680 \text{ watts}\end{aligned}$$

$$\begin{aligned}P_o &= \frac{\mathcal{J}'_a \mathcal{V}'_a}{2} \\ &= \frac{.295 \times 3,600}{2} \\ &= 532 \text{ watts.}\end{aligned}$$

$$P'_d = 148 \text{ watts,}$$

i.e. four times the dissipation under carrier conditions. The output P_o is also equal to $4P_o$ within the limits of the arithmetical accuracy. Thus the modulator will be practically distortionless if the bias is allowed to vary between -365 and -270 volts during the period in which the audio-frequency voltage is additive to the supply voltage. During the other half-cycle the grid will run appreciably more negative than 365 volts, but the operating angle becomes vanishingly small at the negative peak of the modulation cycle; consequently the peak current may theoretically have any value whatever, and the average current will still fall to practically zero. As a result, the variation of I_{av} is nearly symmetrical about the carrier value and no appreciable fluctuation of mean anode current will be observed if the operating conditions are correct.

Grid driving power and bias variation

128. The grid driving power for each of the two above conditions may be found as in the W/T amplifier. Under carrier conditions it is seen from the $I_g - V_a$ curves that $I_{gp} = 125$ milliamperes. As $\cos \psi = \frac{E_b}{V_g} = \frac{365}{565} = .645$, $\psi = 50^\circ$, and reference to fig. 57 shows that $\alpha_g = .143$, so that $I_g = .143 \times 125 = 17.8$ milliamperes. The grid driving power will be

$$\begin{aligned} P_g &= I_g V_g \\ &= .0178 \times 565 \\ &= 10.1 \text{ watts.} \end{aligned}$$

At the peak of the audio-frequency cycle, $I'_{gp} = 75$ milliamperes, and $\cos \psi = \frac{270}{565} = .477$, $\psi = 62^\circ$. Hence $\alpha'_g = .18$ and $I'_g = .18 \times 75 = 13.5$ milliamperes. The grid driving power will therefore be only

$$\begin{aligned} P'_g &= I'_g V_g \\ &= .0135 \times 565 \\ &= 7.7 \text{ watts.} \end{aligned}$$

If the grid bias is to be derived from a condenser with a leak resistance of R_g ohms, we have

$$\begin{aligned} E'_g &= R_g I'_g = 270 \text{ volts} \\ E_g &= R_g I_g = 365 \text{ volts,} \end{aligned}$$

and it is of interest to verify whether these conditions are both satisfied. We have

$$\begin{aligned} R_g &= \frac{270}{I'_g} \\ &= \frac{270}{.0135} \\ &= 20,000 \text{ ohms,} \end{aligned}$$

and also

$$\begin{aligned} R_g &= \frac{365}{I_g} \\ &= \frac{365}{.0178} \\ &= 20,600 \text{ ohms.} \end{aligned}$$

The values of R_g so derived agree to well within the accuracy with which the peak values of grid current can be read off the curves. This is not always so, and it may be found desirable to use a combination consisting of battery (or automatic) grid bias, which is of constant magnitude

CHAPTER XI.—PARAS. 129-131

throughout the audio-frequency cycle, together with a smaller bias voltage derived from a condenser and leak resistance. Such expedients are only necessary where a high degree of fidelity is of primary importance.

The grid-modulated amplifier

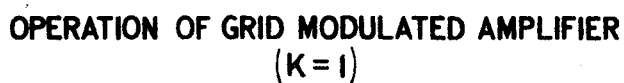
129. In the grid-modulated amplifier the grid-filament P.D. consists of (i) the grid bias voltage, (ii) the radio-frequency excitation derived from the previous stage, e.g. a master oscillator or buffer amplifier, and (iii) an audio-frequency voltage derived from the secondary winding of the microphone transformer, either directly or by means of a sub-modulator valve. The latter voltage, in effect, causes a cyclical variation of grid bias. The operation of this form of modulator is discussed briefly in Chapter XII; it is here only proposed to deal with the possibility of an approach to distortionless modulation, and the power relations. As before, a concrete example will be taken, and fig. 59 shows the ideal characteristics of a valve having the following constants, namely, $r_a = 5,000$ ohms, $\mu = 10$, $P_L = 60$ watts, $I_e = 450$ milliamperes, V_a (max.) = 1,200 volts. The chain-dotted line represents the lower limit of the anode-filament P.D. assuming that it must not be less than the grid-filament P.D. Whereas in the anode-modulated amplifier the operating angle is maintained very nearly constant, grid bias modulation necessitates a cyclical variation of operating angle as the actual bias changes. For 100 per cent. modulation this variation is from 0° to 90° , and when only the radio-frequency excitation is applied, i.e. under carrier conditions, the operating angle is generally very near to 60° .

130. Let us now investigate the conditions under which 100 per cent. modulation may be obtained. Under carrier conditions, the grid will be biased to a value considerably more negative

than the cut-off voltage $\frac{E_a}{\mu}$, the exact value being as yet unknown. At the positive peak of the audio-frequency cycle the grid-filament P.D. may be carried exactly to "cut-off." If then we find the greatest permissible values of I_p and E_g , the required radio-frequency excitation is easily found. These values of I_p and E_g are I'_p and E'_g respectively. The current I'_p is found as in earlier discussions, i.e. $I'_p = .8 I_e = .8 \times 450 = 360$ milliamperes. At the instant when the anode current reaches this value $V_g = E'_g$ and $V_a = E'_o = E'_g$. For ease of location, we may therefore let $I'_p = 350$ milliamperes, and the upper end of the line representing the radio-frequency excitation—the $\mathcal{V}_g$ -line—will be at the point a' , where $E'_g = E'_o = 160$ volts. As the grid bias is to be equal to $\frac{E_a}{\mu}$ at the instant of peak modulation, the operating angle ϕ will be 90° , so that the other end of the $\mathcal{V}_g$ -line will be at the point b .

131. The following data regarding the conditions at the peak of the modulation cycle are now known.

$$\begin{aligned} I'_p &= 350 \text{ milliamperes} \\ \phi' &= 90^\circ \\ \left. \begin{aligned} \alpha' &= .3184 \\ \beta' &= .5 \end{aligned} \right\} &\text{from fig. 57} \\ E'_o &= 160 \text{ volts} \\ E'_g &= 160 \text{ volts} \\ E'_b &= \frac{E_a}{\mu} = 120 \text{ volts} \\ \mathcal{V}'_a &= E_a - E'_o = 1,040 \text{ volts} \\ \mathcal{I}'_a &= \beta' I_p = 175 \text{ milliamperes} \\ I'_{av} &= \alpha' I_p = 110 \text{ milliamperes} \\ \mathcal{V}_g &= E'_g + E'_b = 280 \text{ volts.} \end{aligned}$$



From these we may calculate the instantaneous power input P'_i , output P'_o and efficiency η' . Before doing so, however, the corresponding data will be found for the carrier condition. For 100 per cent. distortionless modulation, we must have

$$\mathcal{I}_a = \frac{\mathcal{I}'_a}{2} = 87.5 \text{ milliamperes}$$

$$\mathcal{V}_a = \frac{\mathcal{V}'_a}{2} = 520 \text{ volts}$$

$$E_o = E_a - \mathcal{V}_a = 680 \text{ volts,}$$

and φ must be about 60° . The upper end of the $\mathcal{V}_g$ -line must lie somewhere upon the vertical through E_o , and its lower end somewhere upon the vertical through E_a . The amplitude $\mathcal{V}_g$ of the radio-frequency voltage is still 280 volts. From these considerations the trial $\mathcal{V}_g$ -line cd was drawn, intersecting the anode voltage base line at $V_a = 930$ volts. We then have

$$\mathcal{V}_a \cos \varphi = 1,200 - 930 = 270$$

$$\cos \varphi = \frac{270}{520} = .52$$

$$\varphi = 58\frac{1}{2}^\circ$$

$$\beta = .36.$$

At the point C , we find $I_p = 215$ milliamperes ;

$$\mathcal{I}_a = \beta I_p = 77.5 \text{ milliamperes,}$$

which is too small. Similarly the $\mathcal{V}_g$ -line ef intersects the base line at $V_a = 1,000$ volts.

$$\mathcal{V}_a \cos \varphi = 200$$

$$\cos \varphi = \frac{200}{520} = .384$$

$$\varphi = 67\frac{1}{2}^\circ$$

$$\beta = .41$$

$$I_p = 275 \text{ milliamperes}$$

$$\mathcal{I}_a = 116 \text{ milliamperes,}$$

which is too large. Now take an intermediate such as $a \ b$, intersecting the base line at $V_a = 950$ volts.

$$\mathcal{V}_a \cos \varphi = 250$$

$$\cos \varphi = \frac{250}{520} = .48$$

$$\varphi = 61^\circ$$

$$\beta = .375$$

$$I_p = 235 \text{ milliamperes}$$

Also $\alpha = .215$ and $I_{av} = 51$ milliamperes.

On this line

$$\begin{aligned} \mathcal{I}_a &= \beta I_p \\ &= 88 \text{ milliamperes,} \end{aligned}$$

which is practically the desired value. Hence we see that the mean bias must be that given by the intersection of this line and the vertical through E_a , and is — 230 volts, so that the amplitude

CHAPTER XI.—PARA. 132

of the audio-frequency grid-filament voltage must be $230 - 120 = 110$ volts. At the negative peak of its cycle, the grid bias will swing down to -340 volts, and as the excitation voltage is 280 , the grid will reach -60 volts. Provided, however, that the output circuit has a high magnification, i.e. low damping, the oscillatory load P.D. will swing almost equally above and below its carrier value, and consequently the operating angle at the point under consideration will be very small indeed. Since for small values of φ , both β and α also become very small, it may be taken that both $\mathcal{I}_a$ and I_{av} will fall to zero at this instant. The amplitude of the oscillatory component of anode current is shown in fig. 59 as $\mathcal{I}_{osc}$, and the average current during the audio-frequency cycle as I_{av} .

132. The power relations will be as follows.

Peak of modulation cycle :—

$$\begin{aligned} P'_i &= I'_{av} E_a \\ &= .110 \times 1,200 \\ &= 132 \text{ watts} \\ P'_o &= \frac{\mathcal{I}'_a \mathcal{V}'_a}{2} \\ &= 175 \times 520 \\ &= 91 \text{ watts} \\ P'_d &= 132 - 91 \\ &= 41 \text{ watts} \\ \eta' &= \frac{91}{132} = 69 \text{ per cent.} \end{aligned}$$

Unmodulated carrier :—

$$\begin{aligned} P_i &= I_{av} E_a \\ &= .051 \times 1,200 \\ &= 61.2 \text{ watts} \\ P_o &= \frac{\mathcal{I}_a \mathcal{V}_a}{2} \\ &= .088 \times 260 \\ &= 22.8 \text{ watts} \\ P_d &= 61.2 - 22.8 \\ &= 38.4 \text{ watts} \\ \eta &= \frac{22.8}{61.2} \\ &= 37 \text{ per cent.} \end{aligned}$$

It will be observed that the supply current rises and falls in a nearly symmetrical manner about its value I_{av} at the carrier condition. It follows that the average power input during operation is approximately equal to P_i , i.e. about 60 watts. During 100 per cent. sinusoidal modulation, the output power increases by the amount necessary to generate the required side-frequencies, i.e. from about 23 to about 34 watts. The average efficiency will be rather higher than that achieved with the unmodulated carrier, but cannot be expected to exceed 60 per cent. When the depth of modulation is less than 100 per cent., the efficiency is correspondingly decreased.

Grid driving power

133. The grid driving power at the modulation peak is found from the $I_g - V_a$ curves as before, taking the conditions at the audio-frequency peak. The peak grid current I'_{gp} will be about 84 milliamperes, and $\cos \psi = \frac{E'_b}{V'_g} = \frac{120}{280} = .43$, hence $\psi \doteq 65^\circ$. Taking $\alpha'_g = .185$ from the square law curve of fig. 57, $I'_g = .185 \times 84 = 15.5$ milliamperes, and

$$\begin{aligned} P_g &= I'_g V'_g \\ &= .0155 \times 280 \\ &= 4.35 \text{ watts.} \end{aligned}$$

It must be particularly noted that this power is necessarily supplied by the audio-frequency source. The average power is of course somewhat smaller than this, but the calculation shows that the microphone transformer, or sub-modulator valve as the case may be, must be capable of delivering a peak current of 84 milliamperes, and yet not cause appreciable distortion during the majority of the cycle when it is very lightly loaded. A permanent resistance shunt across the audio-frequency input terminals, although increasing the load on the source, will tend to reduce distortion.

Automatic grid bias

134. The mean grid bias of a grid-modulated amplifier is often obtained by means of a resistance R_b in series with the negative H.T. supply lead. If this resistance is to carry only the anode current of the amplifier itself, its value is easily calculated. Continuing the previous example, we have found that the mean bias should be 230 volts. As the average direct current under carrier conditions is 51 milliamperes, we have

$$\begin{aligned} E_b &= R_b I_{av} \\ R_b &= \frac{230}{.051} \\ &= 4,500 \text{ ohms.} \end{aligned}$$

Unless certain precautions are taken, the bias will vary during the modulation cycle, e.g. at the positive peak it will be $R_b I'_{av} = 4,500 \times .175 = 790$ volts, and at the negative peak will be zero. It is therefore necessary to shunt the resistance by a condenser of such a value that its reactance X_c at the lowest audio-frequency is small compared to R_b . Suppose we require to transmit audio frequencies down to 200 cycles per second ($\omega = 1,250$), and X_c is to be not more than $.1 R_b$ at this frequency.

$$\begin{aligned} X_c &= \frac{1}{\omega C} = .1 R_b \\ \frac{1}{\omega C} &= 450 \\ C &= \frac{1}{450 \times 1,250} \text{ farad} \\ &= 1.8 \text{ microfarad.} \end{aligned}$$

If, however, the resistance is called upon to carry the anode current of other stages, the variation of grid bias is not so pronounced. Suppose the transmitter to consist only of a master-oscillator

CHAPTER XI.—PARA. 135

and an amplifying stage as above, and the master-oscillator to require a mean anode current of 50 milliamperes. The total current flowing through R_b being denoted by I , we have

At the positive modulation peak $I = .175 + .05 = .225$ amperes.

At the carrier $I = .050 + .050 = .1$ ampere.

At the negative modulation peak $I = 0 + .050 = .05$ ampere,

and the required bias resistance will be approximately 2,300 ohms; if no shunt condenser is fitted the bias will then vary from 520 volts at the positive to 115 volts at the negative peak. A shunt condenser will of course reduce this variation to an inconsiderable amount, nevertheless, where possible battery bias is preferable to automatic bias.

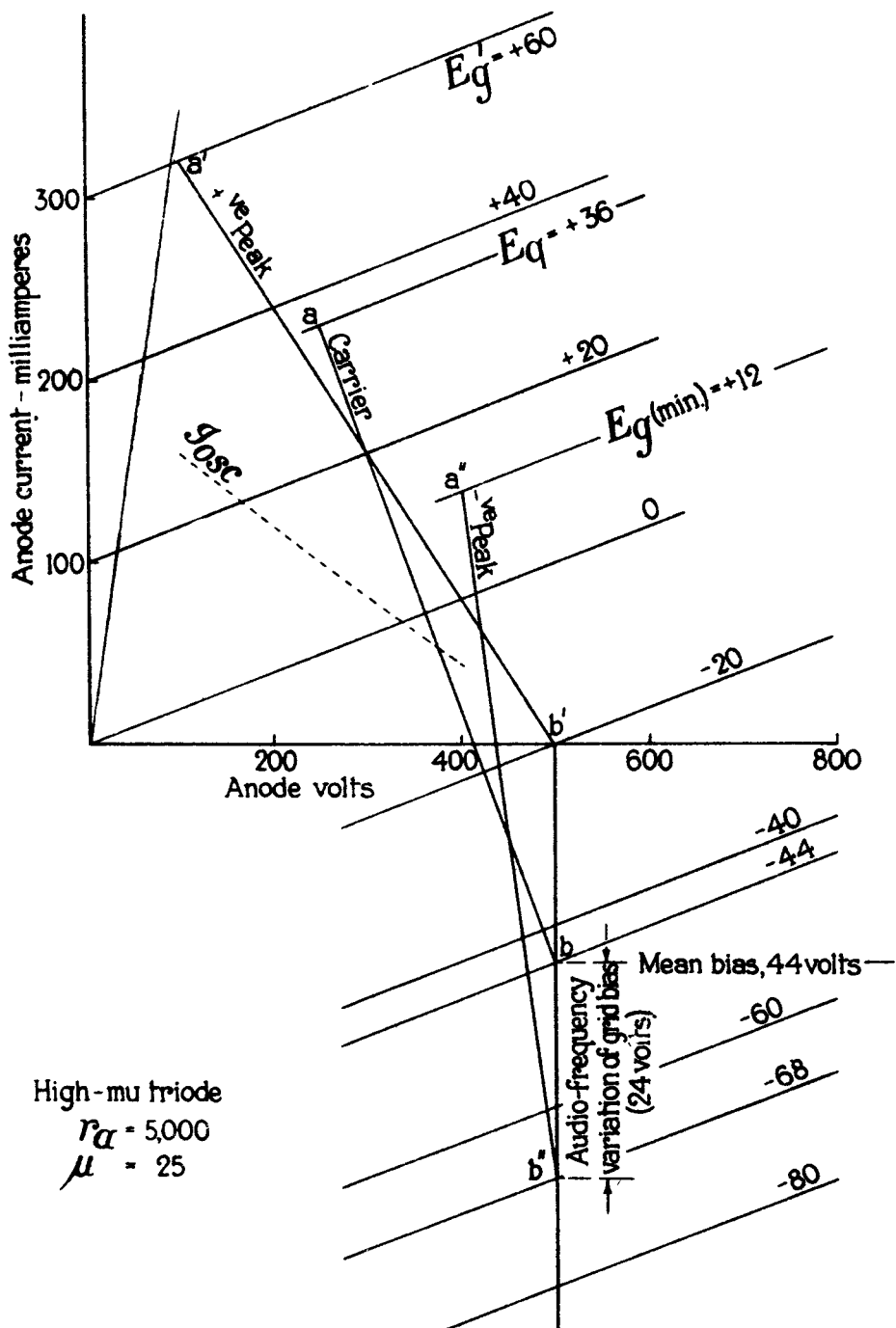
High-mu triode as grid-modulated amplifier

135. In recent years, a coated-filament type of power triode has been developed, ratings up to about 30 watts being available, and these are coming into use in the output stages of low power transmitters. Since with this construction it is possible to obtain a large mutual conductance, a high amplification factor may be obtained with a fairly low anode A.C. resistance; such valves are sometimes referred to as high- μ (i.e. high mutual conductance) triodes. Owing partly to the comparatively small clearance between the electrodes, and partly to the tendency of the coating to disintegrate under a high electrical stress, the maximum permissible H.T. voltage is usually not more than 600 volts. The ideal characteristics of a valve of this type will be used to illustrate the operation of the grid-modulated amplifier under the conditions giving a maximum depth of modulation of the order of 60 per cent. The data for this valve are as follows

$$\begin{aligned} I_e &= 400 \text{ milliamperes} \\ P_L &= 20 \text{ watts} \\ E_a \text{ (max.)} &= 500 \text{ volts} \\ r_a &= 5,000 \text{ ohms} \\ \mu &= 25 \\ \frac{\mu^2}{r_a} &= .125 \end{aligned}$$

At the present stage of development these valves are somewhat prone to trouble due to secondary emission, if the anode-filament P.D. is allowed to fall too low. In the diagram (fig. 60) the chain-dotted line is the limit for E_o on the assumption that E_g is not to exceed $1.5 E_o$. Then for $K = .6$, the $\mathcal{V}_g$ -lines shown have been derived thus. Since $I_e = 400$ milliamperes we have, at the positive modulation peak, $I'_p = .8 I_e = 320$ milliamperes. The point a' on the " +ve peak " $\mathcal{V}_g$ -line $a' b'$ has been located at $I'_p = 320$, $E'_g = 60$, rounding off E'_o to 100 volts instead of 90, and $\mathcal{V}'_a = E_a - E'_o = 400$ volts. For this condition, $\phi = 90^\circ$ and the grid bias $E'_b = -20$, so that a radio-frequency excitation of $\mathcal{V}_g = 80$ volts is required. By methods previously explained it is easily found that $P'_i = 51$ watts, $P'_o = 32$ watts, $P'_d = 19$ watts, $\mathcal{I}'_a = 160$ milliamperes. Under carrier conditions, for 60 per cent. modulation we must have

$$\begin{aligned} \mathcal{I}_a &= \frac{\mathcal{I}'_a}{1 + K} = \frac{\mathcal{I}'_a}{1.6} = 100 \text{ milliamperes} \\ P_o &= \frac{P'_o}{(1 + K)^2} = 12.5 \text{ watts} \\ \mathcal{V}_a &= \frac{\mathcal{V}'_a}{1.6} = 250 \text{ volts.} \end{aligned}$$



OPERATION OF GRID MODULATED AMPLIFIER
 ($K = 0.6$)

FIG. 60
 CHAP. XI

After a few trials, the "carrier" $\mathcal{V}_g$ -line a b was located, giving $I_p = 230$ milliamperes, $\varphi = 71^\circ$, $\alpha = \cdot 25$, $\beta = \cdot 43$. $\mathcal{I}_a = \cdot 43 \times 230 = 99$ milliamperes, which is quite near the required value. Then $P_i = 28\cdot 8$ watts, $P_o = 12\cdot 4$ watts, $P_d = 16\cdot 4$ watts and $\eta = 43$ per cent. On the "—ve peak" $\mathcal{V}_g$ -line a'' b'' we have I_p (min.) = 140 milliamperes, $\varphi = 50^\circ$, $\alpha = \cdot 18$, $\beta = \cdot 31$, $\mathcal{I}_a$ (min.) = 43 milliamperes, rather higher than the required value. The variation of $\mathcal{I}_a$ during the audio-frequency cycle is shown in dotted line as $\mathcal{I}_{osc}$; it is to all intents and purposes linear. It is seen that an audio-frequency variation of grid bias of 24 volts peak will be required, with a mean bias of — 44 volts. This variation may be obtained directly from the secondary of a microphone transformer, without the aid of a sub-modulator stage. The correct load impedance will be $R_d = \frac{\mathcal{V}_a}{\mathcal{I}_a} = 2,500$ ohms.

Neutralization

136. In the operation of radio-frequency power amplifiers it is essential to ensure that no energy is transferred from the output to the input circuit, otherwise the amplifier may become unstable, tending to produce self-oscillation as described in para. 65. Such transfer of energy may be prevented by arranging the circuit in such a manner that an amount of negative reaction

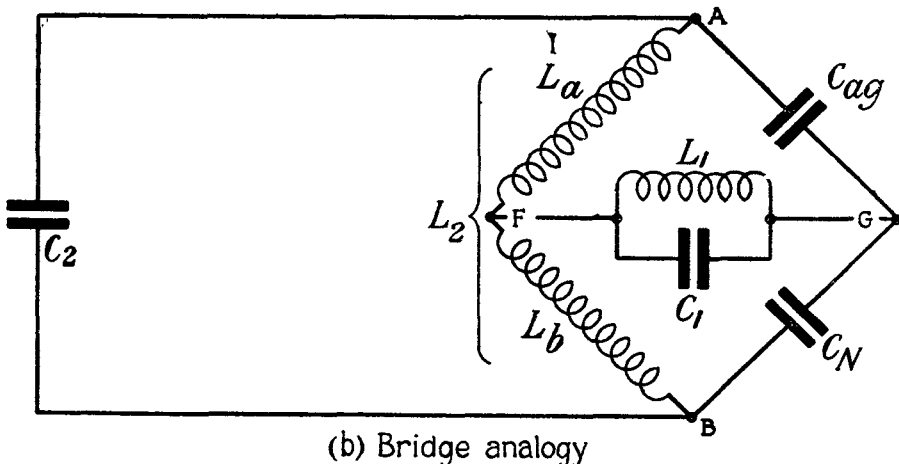
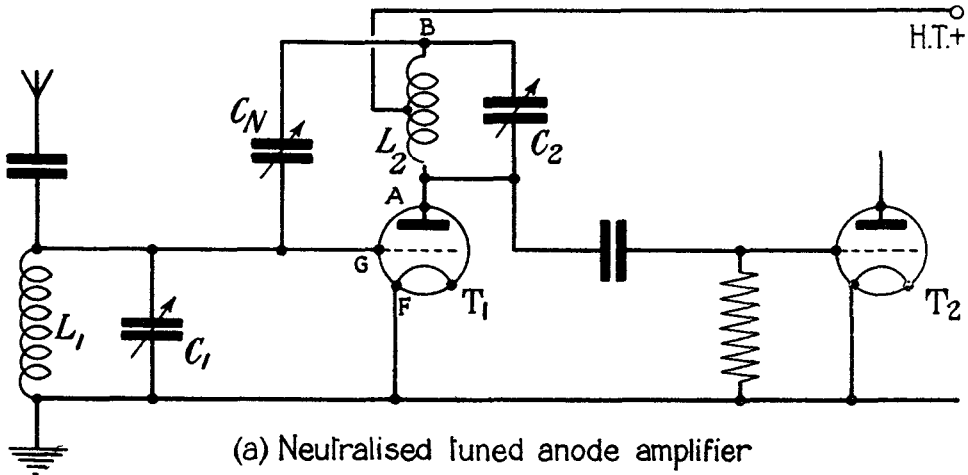


FIG. 61, CHAP. XI.—Principle of neutralization.

CHAPTER XI.—PARAS. 137–138

is introduced, the magnitude of which is exactly equal to the positive reaction which causes the instability. A truly neutralized circuit operates equally well in the opposite sense, i.e. if the valve tends to transfer energy from the input to the output circuit the neutralizing device will apply positive reaction of such magnitude as to prevent this energy transference. Thus, when a valve is truly neutralized, its input admittance is purely capacitive and neither increases nor decreases the damping of the input circuit. This principle was first applied in triode receiving amplifiers, but the introduction of the screen-grid valve has caused its entire disappearance in this field. Nevertheless, it is convenient to refer to this application, and a typical circuit is given in fig. 61 in which a tuned aerial circuit $L_1 C_1$ supplies a signal voltage to the grid and filament of the triode T_1 . The output circuit of the latter is a tuned anode $L_2 C_2$, the positive H.T. supply being connected to a centre tap on the coil L_2 . Provided that the capacitance to earth of the two ends of the coil is the same, equal and opposite P.D.'s must be developed between opposite ends of the inductance and the earthed filament. The grid-anode capacitance C_{ag} is then balanced by connecting a neutralizing condenser C_N between the upper end (B) of the inductance L_2 and the grid. Since the ends (A, B) of the coil are at equal and opposite potentials with respect to the filament, transference of energy between input and output circuits will be entirely prevented by making the capacitance C_N equal to C_{ag} .

137. Provided that the grid-anode admittance is purely susceptive, the equivalent circuit of the amplifier is analogous to the wheatstone bridge described in Chapter I, and the circuit of fig. 60a has been re-drawn (fig. 60b) in order to show the analogue. It is assumed that the P.D. to be neutralized is that of the anode tuning condenser C_2 , which is therefore regarded as the supply voltage to the bridge. The inductance L_2 is split into two portions $L_a L_b$, the midpoint being connected to the filament F, so far as oscillatory currents are concerned. The neutralizing condenser C_N and the grid-anode capacitance C_{ag} are in series between the points A and B, the centre point being connected to the grid G. The input circuit $L_1 C_1$ is connected between G and F and from the bridge point of view may be regarded as analogous to the galvanometer in fig. 16, Chapter I. It is easily seen that the points F and G will be at the same potential if

$$\frac{L_a}{L_b} = \frac{C_N}{C_g},$$

i.e. the balance is independent of frequency. Unfortunately, however, this ideal state of affairs rarely exists in practice, owing chiefly to the presence of conductive paths in parallel with the various reactances. The effect of these shunts is to render the effective values of L_a , L_b , C_{ag} , C_N , dependent upon the frequency, so that it becomes necessary to re-neutralize whenever the frequency is changed appreciably. In a receiving amplifier, however, it is highly desirable to keep the number of controls as low as possible and before the introduction of the screen-grid valve a great deal of ingenuity was expended in the endeavour to maintain a constant neutralizing adjustment over a wide frequency range.

138. In the application of this principle to radio-frequency power amplifiers, it may be regarded as almost axiomatic that complete neutralization over a very wide frequency band is only practicable where weight and space are of no account whatever. These circumstances, however, are only applicable to large ground station installations, where the transmitters are generally operated on a spot frequency for very long periods. On the infrequent occasion of a change of operating frequency, the little extra time required for re-neutralization is of no importance, and there is no necessity to introduce extra complication in an endeavour to maintain a single neutralization adjustment over a wide frequency band. Aircraft transmitters are usually adjusted to a few spot frequencies on the ground (using an artificial aerial) and the tuning and neutralizing adjustments tabulated. Only a slight adjustment of amplifier tuning and neutralization is then required when the aircraft is airborne.

Screen-grid power amplifiers

139. In recent years the screen-grid valve has been developed for use as a radio-frequency power amplifier, both for linear or class B amplification, and as a class C amplifier in W/T transmitters. The screening of this valve is less complete than in the receiving type, and the grid-anode capacitance is by no means negligible, although much smaller than that of the corresponding power triode. The screen-grid power valve is chiefly used in very high frequency transmitters, where it is difficult to obtain an accurate neutralizing adjustment with the power triode. It is usually necessary to incorporate the usual neutralizing arrangements, but it is often found possible to obtain a setting of the neutralizing condenser which will maintain a balance over a very wide frequency range.

Frequency multipliers

140. A valve frequency multiplier is essentially an amplifier operating under conditions which lead to a high degree of amplitude distortion, so that the output contains harmonics of the frequency applied to the grid. The distortion may be produced by the curvature of either the $I_g - V_g$ or the $I_a - V_g$ characteristics, the former being generally used when low power is required with a high multiplication, e.g. for the purpose of radio-frequency measurements. In high and very high frequency transmitters it is usual to attain the desired multiplication by a number of frequency doubling or trebling stages, and the anode current curvature is utilized. The amplitude of any harmonic higher than the third is generally too small to be employed economically. The anode current wave-form of an ideal class B amplifier contains only even harmonics, but the output of a class C amplifier contains both even and odd harmonics, and this type is generally used. The anode circuit is of course tuned to the harmonic frequency and not to the frequency of the input, consequently the circuit has no tendency to self-oscillation and does not require neutralization. If the $I_a - V_a$ characteristics of the valve are available, its performance as a frequency doubler or tripler can be calculated in a manner similar to that adopted when its output circuit is tuned to the input frequency, the curves shown in dotted line in fig. 57 being used for this purpose. They give the ratios

$$\beta_2 = \frac{\mathcal{J}_2}{I_p}$$

and

$$\beta_3 = \frac{\mathcal{J}_3}{I_p}$$

where $\mathcal{J}_2$ and $\mathcal{J}_3$ are the amplitudes of the second and third harmonics, respectively, of the fundamental component of anode current. The desired harmonic is of course selected by suitably tuning the anode circuit, and the load P.D. V_a will be, to all intents and purposes, sinusoidal and of the selected harmonic frequency.

141. Before giving typical numerical examples the general trend of these curves should be observed. Take the curve $H_2(1)$ giving the values of β_2 for first-power-law characteristics.

The output power is equal to $\frac{\beta_2}{2} I_p (E_a - E_0)$ and the input to $a I_p E_a$. It follows that, for a

given ratio of E_0 to E_a , the efficiency is directly proportional to $\frac{\beta_2}{a}$, and the product ηP_0

directly proportional to $\frac{\beta_2^3}{a}$. By actual point-to-point calculation for various values of φ_2 , it is

found that for large output and high efficiency the operating angle should be between 40° and 70° , maximum output being achieved when $\varphi_2 \doteq 65^\circ$. The efficiency increases as φ_2 is reduced and ηP_0 is a maximum when $\varphi_2 \doteq 60^\circ$. Similar considerations may be applied to the curve giving the values of β_3 against φ_3 , which is applicable to the frequency trebler; in this case, the output

CHAPTER XI.—PARA. 142

is a maximum (for first-power-law characteristics) when $\varphi_3 \doteq 40^\circ$, and the product ηP_o a maximum with a rather smaller operating angle. It is seen then that where β_2 or β_3 is very nearly constant over a range of values of φ , the smaller operating angle is preferable in order to reduce the power dissipation and increase the efficiency. It will however be found that a small operating angle necessitates a large grid excitation, and a consequent increase in grid driving power.

Power output and efficiency of frequency doubler

142. To illustrate the method of calculating the power relations, we shall take the ideal characteristics of fig. 62, the permissible dissipation being 250 watts and the maximum supply voltage 5,000 volts. Let $E_a = 4,000$ volts, $E_o = E_g = 300$ volts, and $I_p = 630$ milliamperes; if the valve is to operate as a frequency doubler, reference to the curve H_2 (1) of fig. 57 shows that the maximum value of β_2 is achieved with an operating angle $\varphi_2 = 65^\circ$. We then have

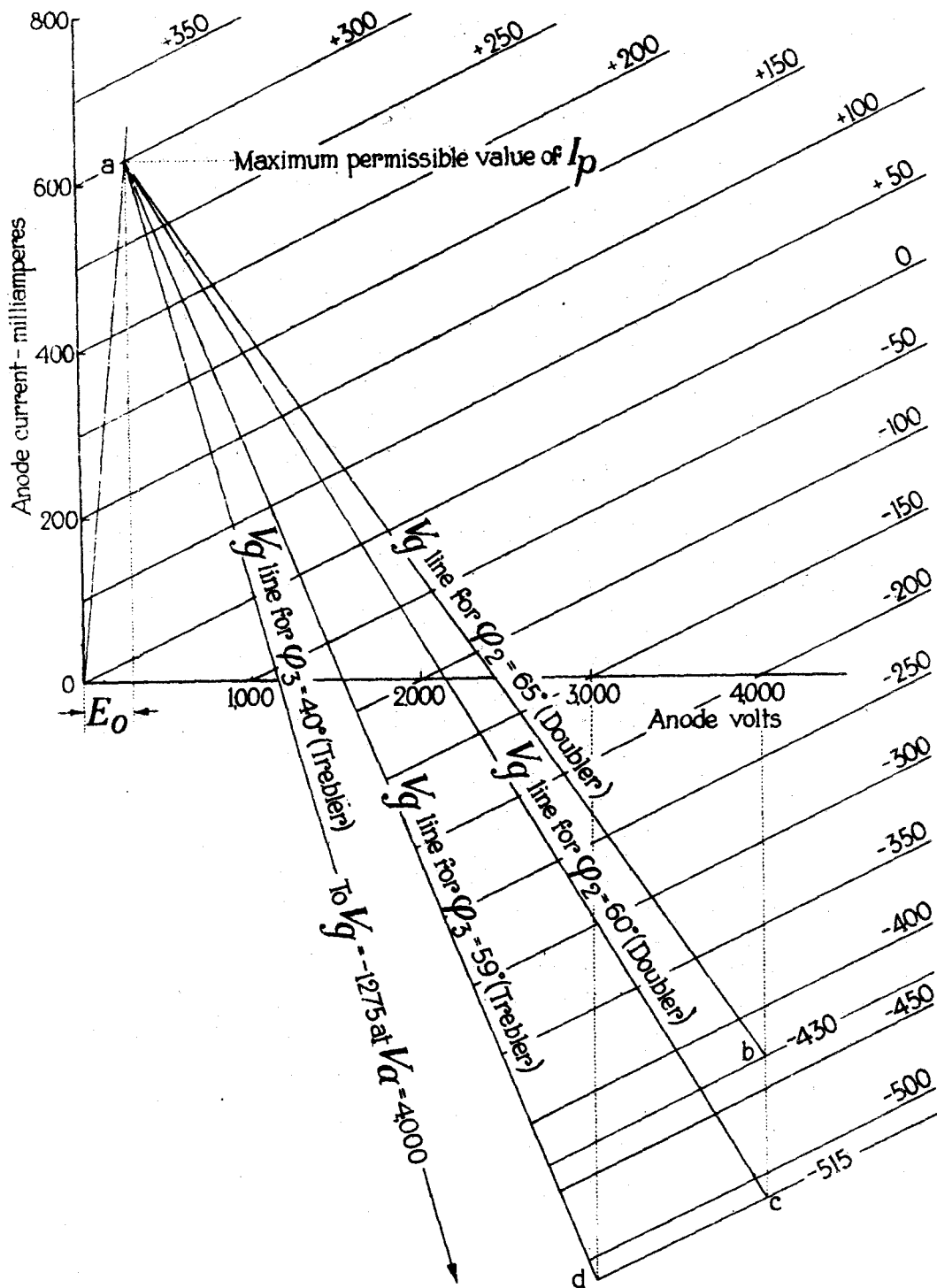
$$\begin{aligned}\mathcal{V}_a &= E_a - E_o = 3,700 \text{ volts} \\ \varphi_2 &= 65^\circ \\ \cos \varphi_2 &= .423 \\ \mathcal{V}_a \cos \varphi_2 &= .423 \times 3,700 \\ &= 1,560 \text{ volts.}\end{aligned}$$

The $\mathcal{V}_g$ -line a b has been drawn through the anode voltage base line at $4,000 - 1,560 = 2,440$ volts, intersecting the vertical through E_a at a point giving $E_b = 430$ volts. Hence the required excitation $\mathcal{V}_g$ is $430 + 300 = 730$ volts. If this can be provided we may proceed to find the power relations.

$$\begin{aligned}P_i &= a I_p E_a \\ &= .23 \times .63 \times 4,000 \\ &= 580 \text{ watts} \\ P_o &= \frac{\beta_2}{2} I_p \mathcal{V}_a \\ &= \frac{.275 \times .63 \times 3,700}{2} \\ &= 322 \text{ watts} \\ P_d &= 258 \text{ watts} \\ \eta &= 55 \text{ per cent.}\end{aligned}$$

Repeating the calculation for an operating angle $\varphi_2 = 60^\circ$, we find

$$\begin{aligned}\mathcal{V}_a &= 3,700 \\ \cos \varphi_2 &= .5 \\ \beta_2 &= .274 \\ a &= .212 \\ P_i &= .212 \times .63 \times 4,000 \\ &= 535 \text{ watts} \\ P_o &= \frac{.274 \times .63 \times 3,700}{2} \\ &= 320 \text{ watts.} \\ P_d &= 215 \text{ watts} \\ \eta &= 60 \text{ per cent.}\end{aligned}$$



VARIOUS OPERATING CONDITIONS FOR FREQUENCY
MULTIPLICATION

The required bias and excitation will be derived algebraically. From paragraph 120,

$$\begin{aligned} E_b &= \frac{E_a}{\mu} + \left(E_g + \frac{E_o}{\mu} \right) \frac{\cos \varphi_2}{1 - \cos \varphi_2} \\ &= 200 + (300 + 15) \frac{.5}{.5} \\ &= 515 \text{ volts} \end{aligned}$$

and the excitation

$$\mathcal{V}_g = E_b + E_g = 815 \text{ volts.}$$

The $\mathcal{V}_g$ -line is therefore a c.

Power output and efficiency of frequency trebler

143. Now suppose the same valve to be operated as a frequency trebler. Reference to curve H_3 (1) shows that β_3 attains its greatest value if $\varphi_3 \doteq 40^\circ$. If $E_a = 4,000$, $E_o = E_g = 300$ as before, $\mathcal{V}_a = 3,700$ and

$$\mathcal{V}_a \cos \varphi_3 = 2,840$$

The $\mathcal{V}_g$ -line from the point a will therefore intersect the anode base-line at $V_a = 1,160$ volts. It will intersect the vertical through $E_a \doteq 4,000$ at a point corresponding to $V_g = -1,275$ volts (approximately) and an excitation $\mathcal{V}_g = 1,575$ volts will be necessary. It may be impracticable to provide such a large excitation, but for purposes of comparison the power relations are

$$\begin{aligned} P_i &= \alpha I_p E_a \\ &= .145 \times .63 \times 4,000 \\ &= 366 \text{ watts} \\ P_o &= \frac{\beta_3}{2} I_p \mathcal{V}_a \\ &= \frac{.185 \times .63 \times 3,700}{2} \\ &= 216 \text{ watts} \\ P_d &= 150 \text{ watts} \\ \eta &= 59 \text{ per cent.} \end{aligned}$$

From the equation

$$E_b = \frac{E_a}{\mu} + \left(E_g + \frac{E_o}{\mu} \right) n$$

where

$$n = \frac{\cos \varphi_3}{1 - \cos \varphi_3}$$

it is easy to find the excitation required for any given values of E_a , E_g , E_o and n . If we still intend to try to operate with $\varphi_3 = 40^\circ$ ($n = 3.4$) it is found that since $\left(E_g + \frac{E_o}{\mu} \right) n = 3.4 \times 315 = 1070$ and $\mathcal{V}_g = E_b + E_g$

$$\begin{aligned} E_b &= \frac{E_a}{20} + 1,070 - 300 \\ &= \frac{E_a}{20} + 770, \end{aligned}$$

e.g. if $E_a = 1,000$, we require a bias of 820 volts and an excitation of 1,120 volts to give the desired operating angle. The output would be very small, of the order of only 60 watts for an

CHAPTER XI.—PARA. 144

input of 90 watts. The efficiency is however fairly high. Alternatively we may try a less drastic reduction of E_a , and operate with an angle greater than the optimum. Thus if $E_a = 3,000$ and $E_b = 515$ volts, the $\mathcal{V}_g$ -line will be a d, fig. 62.

$$\begin{aligned} n &= \frac{E_b - \frac{E_a}{\mu}}{E_g + \frac{E_o}{\mu}} \\ &= \frac{515 - 150}{315} \\ &= 1.06 \end{aligned}$$

also

$$\begin{aligned} \phi_3 &= \frac{n}{n + 1} \\ &= \frac{1.06}{2.06} \\ &= .514 \\ \phi_3 &= 59^\circ. \end{aligned}$$

Then

$$\begin{aligned} \beta_3 &= .146 \text{ (from fig. 57)} \\ \mathcal{V}_a &= 2,700 \\ P_o &= \frac{.146 \times .63 \times 2,700}{2} \\ &= 125 \text{ watts} \\ P_i &= .21 \times .63 \times 3,000 \\ &= 400 \text{ watts} \\ P_d &= 275 \text{ watts} \\ \eta &= 31 \text{ per cent.} \end{aligned}$$

The dissipation is now excessive and the efficiency low.

144. If the characteristics are considerably curved, so that the harmonics are more prominent, the performance may be rather better than the foregoing figures suggest. Thus if we estimate the output of the previous example from the H_3 (2) curve for β_3 , with the operating angle $\phi_3 = 59^\circ$, we obtain

$$\begin{aligned} \beta_3 &= .179 \\ P_o &= \frac{.179 \times .63 \times 2,700}{2} \\ &= 150 \text{ watts} \\ P_i &= .17 \times .63 \times 3,000 \\ &= 320 \text{ watts} \\ P_d &= 170 \text{ watts} \\ \eta &= 47 \text{ per cent.} \end{aligned}$$

The actual performance almost certainly lies somewhere between the two, and we may conclude that the output will be of the order of 140 watts, the dissipation rather more than the output, and the efficiency rather less than 50 per cent.

CHAPTER XII.—RADIO-TELEPHONY

1. The production of continuous and interrupted continuous waves for radio-telegraphic purposes having been discussed in Chapter IX it is now proposed to consider the principles of radio-telephony. Radio-telephony is the transmission of speech, music or any other form of sound by means of electro-magnetic waves. As a means of communication radio-telegraphy and radio-telephony may be compared as follows :—

(i) Telephony possesses the advantages of verbal over written communication, i.e. saving of time, and rapid adjustment of minor points of detail. Arrangements are generally possible whereby the communication is carried out directly by the responsible officers, without the intervention of an operator, since knowledge of the morse code is not necessary. This is the most outstanding advantage from the service point of view.

(ii) Its disadvantages, compared with W/T are, first, that no written record of direct communication is available, and if such a record is required, the rapidity of communication is less than by telegraphy. Second, it is difficult to transmit by telephony messages in code, while the use of plain language involves risk of interception. Third, listening-through is not possible except by the use of complicated arrangements, which are impracticable for service use. Fourth, interference is more troublesome than with W/T.

SOUND

2. A brief reference to the phenomena connected with the process of hearing has already been made. (Chapter X.) Sound is invariably caused by the vibration of a material body, and the ensuing undulation of the medium in which the source of sound is situated constitutes a sound wave. In the absence of some such medium, no sound wave is produced, the usual method of demonstrating this being to enclose an electric buzzer in the bell-jar of an air-pump. When the bell-jar contains air at normal atmospheric pressure, the passage of sound waves is only slightly hindered by the presence of the jar, and the buzzer is distinctly heard. As the jar is evacuated, however, the sound becomes progressively fainter, and eventually becomes inaudible. This experiment shows that sound waves are of a nature different from light waves, for the bulb of an ordinary incandescent lamp (i.e. the "vacuum", as opposed to the "gas-filled" type) is evacuated to the highest degree possible, but the radiation of light is unhindered by the absence of a material medium. So far as our normal experience is concerned, the medium in which sound waves are most often propagated is air, but any form of matter, solid, liquid or gaseous, will transmit sound, the motion of the particles in the medium being to and fro in the direction of propagation, in contrast to the corresponding motion in, say, a surface wave in water; it is easily observed that in the latter instance the particle movement is in a plane perpendicular to the direction of propagation of the wave. Surface waves in water are an example of what are called transverse waves, while sound waves exemplify a type of undulation known as longitudinal wave motion.

Characteristics of sound

3. Sounds are conveniently divided into two classes, namely those which produce a pleasing effect upon the ear, and those which are unpleasant. The former are called musical sounds and the latter noises, but it is difficult to draw a definite line of demarcation. A sound wave possesses three characteristic properties by which it can be distinguished. These are :—

- (i) Its wave form, also referred to as its quality or timbre.
- (ii) Its intensity.
- (iii) Its pitch.

The wave form in turn may be broadly divided into two classes, namely, repetitive and non-repetitive, which corresponds to the previous division into musical sounds and noises, and it

CHAPTER XII.—PARAS. 4-5

may be said that a musical sound is that caused by a body whose vibrations possess a repetitive wave form, while noises are produced by bodies in a state of irregular vibration. In telephony, the reproduction of noises and musical sounds are of equal importance. In the service we are mainly concerned in the transmission of speech, which may be said to consist of an irregular succession of both musical and non-musical sounds, the former being the vowel sounds and the latter the consonants. Owing to their sustained character, vowels are generally easily recognizable in telephonic transmission, but the brief duration of the consonant sounds renders it necessary to articulate them with special care (though not with exaggerated emphasis). Noises of brief duration, including the sounds caused by percussion instruments, e.g. drums, and sounds having repetitive but heavily damped wave forms such as those produced by the piano and triangle are often classed together under the term "transients". Complex wave forms consisting of a fundamental and a succession of harmonics have already been considered in Chapter V, but harmonics are of far greater importance in sound than in ordinary A.C. engineering, because the characteristic quality or timbre of any particular instrument depends upon its harmonic content. Thus the tuning fork and the flute emit waves which are almost truly sinusoidal, whereas the violin emits a wave which is particularly rich in harmonics. The wave form is partly governed by the properties of the sounding board upon which the vibrating body is mounted, e.g. the body of the violin, and certain German physicists contend that the timbre of a musical instrument is partly due to the emission of a heavily damped transient which occurs at the beginning of every cycle of the fundamental oscillation, and is termed a "tone-former"

4. The intensity of a sound is a purely physical property independent of the ear, or other receiving device, and is proportional to the square of the amplitude of the wave. In practice the terms intensity and loudness are often regarded as synonymous. Loudness, however, is an indication of the degree of sensation produced in the brain of the hearer, and is difficult to define. The pitch of a sound depends upon the frequency of the disturbance, or in the case of a complex wave form upon the fundamental frequency. It has already been stated that the normal human ear will respond to frequencies ranging from about 16 to 20,000 cycles per second. It is useful to fix a mental standard of pitch, by comparison with the keyboard of a piano. Middle C (a white note immediately to the left of two black notes, a little to the left of the lock) has a frequency which for scientific purposes is taken as 256 cycles per second, but which in modern orchestras is given a higher pitch by increasing the tension on the strings; when tuned to concert pitch, middle C has a frequency of 261.65 cycles per second. The seventh white note above is called first upper C, or C^1 , and has a frequency twice that of middle C, i.e. 512 on the scientific scale or 523.3 cycles per second in concert pitch, while first lower C, or C_1 , has a frequency of 128 and 130.8 on the respective scales. The normal human voice ranges from 80 cycles per second (which is the lowest note usually reached by bass singers) to about 1,200 which is reached by some sopranos. These frequency numbers refer to the fundamental in each case, but harmonics are always present. For good reproduction of human speech in telephony it may be assumed that all frequencies between 200 and 2,000 cycles per second must be retained; frequencies outside this range contribute but little to the intelligibility, although they serve to distinguish the voice of one person from that of another.

The organ of hearing

5. The physiological process which we refer to as "hearing" is only imperfectly understood, but it is generally believed to be principally a phenomenon of resonance. A sound wave gathered by the pinna or external ear enters the ear passage, which is terminated by a membrane commonly though wrongly called the ear drum. The latter is really the cavity at the rear of the membrane. In contact with the membrane is the first of a train of small bones, the last of which communicates with an oval membrane at the other end of the cavity. The real process of hearing appears to commence at this point. The inner structure is very complex; the vibrations which are set up in the outer chamber are communicated to the inner structure, and in particular to the basilar membrane, which consists of a large number of tightly stretched strings like those of a harp, each of which is connected to the brain by a nerve. According to the resonance theory any vibrator reaching the basilar membrane throws into vibration just that portion which is

tuned to the particular frequency, and impulses are transmitted through the nervous system to the brain, giving rise to a sensation of sound of the particular pitch, loudness and duration. It will be noted that the wave form is not conveyed to the brain, for the ear resolves a complex wave into its constituent vibrations. As a result of this analysis, the ear is unable to appreciate the relative phase of the harmonics contained in a complex wave form. For example, the waves shown in heavy line in fig. 1 produce the same aural impression, although their shape is entirely different. Both waves consist of a fundamental frequency and its third harmonic; in fig. 1a the peak value of the harmonic always occurs at an instant when the fundamental is passing through its peak value but of opposite sign. Such a phase relationship gives rise to what is commonly referred to as a flat-topped wave. In fig. 1b, the peak value of the harmonic always occurs at an instant when the fundamental is passing through its peak value of the same sign, giving rise to what is called a peaky wave. The fact that the ear automatically resolves a complex wave form into its constituent frequency components is of considerable importance, in that a sound-reproducing system which does not transmit all frequencies in their correct relative phase may still appear to give distortionless reproduction.

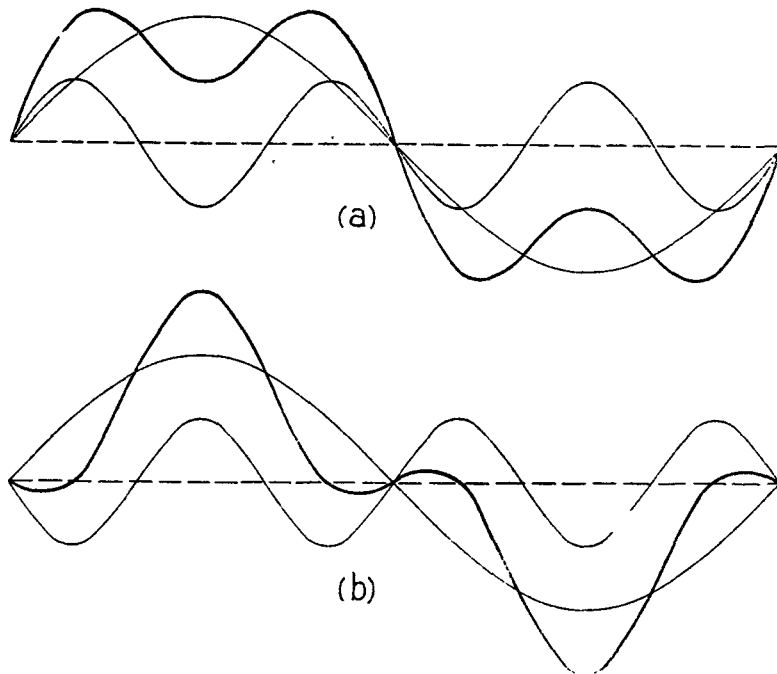


FIG. 1, CHAP. XII.—Waves with positive and negative third harmonic.

MICROPHONES

Pressure and velocity microphones

6. In telephony, the ultimate object is to transmit oral intelligence by electrical means, and the first stage of such a transmission is the conversion of sound waves into electrical impulses of some kind. Instruments designed to this end are called microphones, the original type being the carbon microphone which was briefly described in Chapter I. It will be remembered that this instrument operates by virtue of the change of resistance of the carbon granule pack under the influence of sound waves; other methods of conversion may be utilized, and the forms of microphone which have been developed may be classified in several ways. Before dealing with these, it should be appreciated that in all forms of microphone, the sound wave is first caused to

CHAPTER XII.—PARAS. 7-8

set in vibration a light diaphragm, which in turn initiates some electrical action, which differs in the different types. Since sound waves are of the longitudinal type, the diaphragm itself may be set in motion in two different ways. First suppose the diaphragm to be mounted in such a manner as to form one side of a closed box, the remaining sides being very rigid compared to the diaphragm. A sound wave in air consists of a progressive variation of the pressure of the atmosphere (of the order of about 10 dynes per centimetre for ordinary speech, within a few inches from the mouth), and this variation of pressure will set up a corresponding vibration of the diaphragm, because its inner side is not exposed to the pressure variation owing to the rigidity of the other sides of the box. If, however, the same diaphragm is suspended in the air so that the variation of atmospheric pressure affects both sides simultaneously, it cannot be set in vibration by this means. It is thus apparent that when only one side of the diaphragm is exposed to the pressure variation, the diaphragm is set in vibration no matter in what position it is placed relative to the direction of propagation, for fluid pressure is transmitted equally in all directions. A microphone having a diaphragm which is actuated by the alternate compression and rarefaction of the air is called a pressure microphone. To a first approximation it may be said to be non-directional.

7. (i) Now consider the same diaphragm to be mounted in such a manner that its edge is fixed in space and both faces are exposed to the air; owing to its flexibility the centre portions of the diaphragm may still be set in vibration. Suppose a sound wave to be propagated in the vicinity of the diaphragm, then at any instant, although the air pressure may be above or below normal, both sides are affected simultaneously and to the same degree, hence the diaphragm is not set in vibration by the variation of pressure. There is, however, another phenomenon which must be taken into account. Since the vibration is longitudinal, the air particles are also in vibration to and fro along the line of propagation, and if the diaphragm is so placed that its face is perpendicular to the direction of propagation, it will be set in vibration by the actual movement of the air particles, while if its face is parallel to the direction of propagation no vibration will take place. A microphone in which the diaphragm is actuated in this manner is called a velocity microphone. Its characteristic is that it possesses more or less marked directional properties.

(ii) It is obviously desirable that whatever the mechanism of the microphone itself, the waveform of its ultimate electrical output should be a faithful copy of the waveform of the sound by which it is operated. In radio-telephony, the electrical output is usually required in the form of a voltage, current and power being of subsidiary importance. For our purpose then we may say that the ideal pressure microphone should possess such characteristics that the output voltage V_o is directly proportional to the instantaneous pressure p due to the sound wave, i.e. $V_o = Kp$, and correspondingly, in the velocity microphone the output voltage should be directionally proportional to the instantaneous velocity u of the air particles, i.e. $V_o = Ku$. No type of microphone fulfils the above requirement over the whole range of frequency and amplitude, but certain types give a satisfactory approach to the ideal between certain limits. It is however unfortunate, from the service point of view, that the lines of attack appear invariably to lead to greater weight and linear dimensions, to a less robust instrument, and to reduced sensitivity compared with the carbon microphone. The simplicity and sensitivity of this instrument lead to its almost universal adoption for aircraft R/T installations.

Carbon microphones

8. (i) The carbon microphones in general use may be divided into two classes, namely those designed for use on the ground (such as the hand press and breastplate microphones) and those used in the air (mask microphones). The principles are the same in both classes, the objects of the mask microphone being first, to leave the hands free to perform other duties during the act of transmission, and second, to allow the wearer to assume any position these duties require, e.g. he may be required to use the microphone while in a prone position. The main features of the hand press microphone are shown in fig. 2. It consists of a hollow metal cylinder which is made in two parts in order to give access to the interior. The rear end of the cylinder carries ebonite terminal blocks, to which are attached the ends of the flexible cable by which the

microphone is connected in circuit. These terminal blocks carry german silver springs by which contact with the microphone capsule is maintained. The capsule in turn is of aluminium and carries at its wider end a mica diaphragm often called the sounding board. The microphone button is so mounted that one electrode is rigidly held by the rear end of the capsule but is insulated therefrom by fibre washers, while the other electrode, namely that mounted upon the mica diaphragm of the microphone button, is held on the centre of the sounding board by nuts and washers. This electrode is connected to the metal body of the capsule by means of a flat copper strip; in the diagram, however, this connection is shown as a short piece of insulated wire. It will be seen that the lower spring contact makes directly on a threaded extension of the rear electrode, while the upper spring contact is in metallic connection with the front electrode of the microphone button. A short trumpet-shaped extension of ebonite projects to the front of the casing and gathers the speech waves in such a manner that the actual variation of pressure

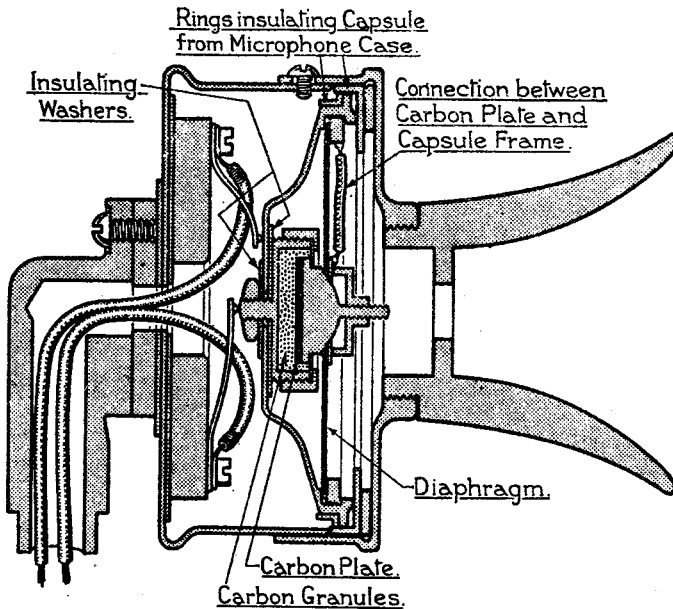


FIG. 2, CHAP. XII.—Head of hand press microphone.

upon the sounding board is greater than the variation of pressure at the open end of the mouth piece. The action of this mouthpiece, in conjunction with the space between its smaller end and the sounding board, is analogous to that of a step-up transformer.

(ii) In the mask microphone no capsule or sounding board is fitted and the button itself differs in design from that shown in fig. 2. Two mica diaphragms are fitted, each carrying a carbon electrode. One diaphragm is mounted on each side of a circular hole in an ebonite plate, the space so enclosed being partly filled with carbon granules. This type of microphone is fully described and illustrated in the appropriate chapter of A.P. 1186, Signal Manual Part IV. With regard to its action, it will be seen that an increase of pressure on both diaphragms simultaneously will lead to a reduction in the resistance of the carbon granule pack, while the reverse effect will be caused by a reduction of pressure. Thus a microphone constructed in this manner is of the pressure type.

9. The mean resistance of both the above forms of microphone button is generally of the order of 50 ohms; but depends to some extent upon the current flowing. It must be borne in mind that if the feed current (i.e. the mean steady current) exceeds about 150 milliamperes,

CHAPTER XII.—PARA. 10

arcing will occur between the sharp edges of adjacent granules and they tend to become welded together; the microphone is then said to be caked. This condition is also sometimes caused by the penetration of moisture into the interior of the button. In any event, it is advisable to give the casing of the microphone a sharp tap with the finger before commencing to speak, in order to shake up the granules and so restore the instrument to its most sensitive state. The use of feed currents higher than those recommended for any particular type will soon cause serious deterioration, because the effect of arcing between the granules is to remove their sharp edges, upon which the sensitivity of the microphone depends.

Audio-frequency voltage generated by carbon microphone

10. (i) The action of a carbon microphone depends upon its variation of resistance, and the latter in turn upon the change of pressure upon the diaphragm due to the impressed sound wave. It is desirable to possess some idea of the magnitude of the effect, although an exact numerical treatment is very difficult. It may, however, be assumed that the diaphragm and granules normally exert a slight mutual pressure P , owing to the weight of the granules and the stiffness of the diaphragm. Let the normal resistance of the microphone be R_m ohms. An increase of pressure then causes a decrease of resistance and vice-versa, so that the following law may be expected to hold, viz. :—

$$R_m = \frac{K}{P}$$

Now consider a varying pressure $p = P_o \sin \omega t$ to be applied to the diaphragm. The resistance will then become $R_m + r_m$ where

$$R_m + r_m = \frac{K}{P + p} = \frac{K}{P \left(1 + \frac{p}{P}\right)}$$

Now

$$\frac{1}{1 + \frac{p}{P}} = 1 - \frac{p}{P} + \left(\frac{p}{P}\right)^2 - \left(\frac{p}{P}\right)^3 \dots\dots\dots,$$

and if the maximum value of p , i.e. P_o , is small compared with the normal pressure P , $\left(\frac{p}{P}\right)^3$ will be much smaller than unity, while higher powers will be absolutely negligible, and

$$\begin{aligned} R_m + r_m &= \frac{K}{P} \left\{ 1 - \frac{p}{P} + \left(\frac{p}{P}\right)^2 \right\} \\ &= R_m \left\{ 1 - \frac{p}{P} + \left(\frac{p}{P}\right)^2 \right\} \end{aligned}$$

or

$$r_m = R_m \left\{ -\frac{p}{P} + \left(\frac{p}{P}\right)^2 \right\}.$$

This then is the magnitude of the variation in resistance of the microphone; experiment shows that the peak pressure upon the diaphragm during ordinary speech is of the order of 30 dynes per square centimetre, while the effective back pressure P of the diaphragm in a certain design was found to be of the order of 500 dynes per square centimetre. Suppose R_m to be 50 ohms, then

$$\begin{aligned} r_m &= 50 \left\{ -\frac{30}{500} + \left(\frac{30}{500}\right)^2 \right\} \\ &= -3 \text{ ohms (nearly)} \end{aligned}$$

The feed current to the microphone in an ordinary series circuit will be

$$I_m = \frac{E}{R + R_m + r_m}$$

where R is the D.C. resistance of the external circuit and E is the applied E.M.F.. As r_m is quite small compared to $R + R_m$, however, I_m may be regarded as being constant and equal to $\frac{E}{R + R_m}$. The P.D. between the terminals of the microphone may be considered to consist of two components, namely a constant voltage $V_m = \frac{R_m E}{R + R_m}$, with which we are not immediately concerned, and a varying component $v_m = r_m I_m$. Then since $p = P_o \sin \omega t$, where $\frac{\omega}{2\pi}$ is the frequency of the sound wave acting on the diaphragm,

$$v_m = I_m R_m \left\{ -\frac{P_o \sin \omega t}{P} + \left(\frac{P_o \sin \omega t}{P} \right)^2 \right\}$$

For example, if $I_m = 150$ milliamperes, $R_m = 50$ ohms, $P_o = 30$ dynes per square centimetre, $P = 500$ dynes per square centimetre, we have

$$\begin{aligned} v_m &= .15 \times 50 (-0.06 \sin \omega t + .0036 \sin^2 \omega t) \\ &= -0.45 \sin \omega t + .027 \sin^2 \omega t. \end{aligned}$$

It will be observed that in obtaining a value for r_m the term $\left(\frac{p}{P}\right)^2$ was ignored, but that it has been retained in deriving an approximate numerical value for the effective voltage v_m . The object of keeping it in the second instance is to show the relative amplitude of the second harmonic of the speech frequency. As $\sin^2 \omega t = \frac{1}{2}(1 - \cos 2\omega t)$, we see that the microphone also generates a second harmonic having an amplitude, in general, equal to $\frac{P_o}{2P}$ times that of the fundamental; in the particular case under consideration, the amplitude will be $\frac{.027}{2} = .0135$ volts. Harmonics of higher order will also be present, but provided that P_o is small compared to P , their amplitude will be inconsiderable.

(ii) Summarizing the above, then, it may be said that under the influence of a sound wave the carbon microphone may be considered to act as an alternating current generator, its E.M.F. waveform having a fundamental and a series of harmonics for each frequency contained in the original speech wave; unless the pressure on the diaphragm due to the sound wave is of the same order as the normal pressure, the harmonics introduced by the microphone itself will be negligible. A more serious consequence of its non-linearity, particularly when used in noisy surroundings, is the phenomenon called intermodulation, which is discussed in paragraph 41.

The condenser microphone

11. (i) This consists of an air dielectric condenser, the capacitance of which is varied by the action of sound waves upon one electrode, which is made in the form of a very thin disc, usually of duralumin. The other electrode is a rigid brass disc, and the thickness of the dielectric is of the order of .002 inch. Thus, if the effective diameter of the electrodes is 1.5 inches, the normal capacitance will be about .0002 μ F. When sound waves impinge upon the diaphragm, its movement results in a variation of the thickness of the dielectric and a corresponding change in capacitance. The manner in which this change is caused to develop a varying voltage is shown in fig. 3. Here E is an E.M.F. set up by a battery of some 200 volts, L an audio-frequency choke of several hundred henries, which however, may be replaced by a large non-inductive resistance in some instances, and C the microphone. Under normal conditions the P.D. across the microphone terminals is equal to the E.M.F. of the battery, i.e. E volts.

Let the normal capacitance of the microphone be $C_o = \frac{K}{D_o}$ where K is a constant and D_o the thickness of the dielectric. When the polarizing voltage E is applied to the condenser, the

CHAPTER XII.—PARA. 11

condenser will receive a charge Q , equal to $C_o E$. Provided the inductance is sufficiently large, this charge will remain sensibly constant even if the capacitance is varied, and as a result we have a varying P.D. at the microphone terminals. Thus if $d = D_o + D \sin \omega t$

$$C = \frac{K}{d} = \frac{K}{D_o (1 + \frac{D}{D_o} \sin \omega t)} = \frac{C_o}{1 + \frac{D}{D_o} \sin \omega t}.$$

Also $Q = Cv = C_o E$ where v is the P.D. at the microphone terminals.

$$\begin{aligned} v &= \frac{C_o}{C} E \\ &= C_o E \frac{\left(1 + \frac{D}{D_o} \sin \omega t\right)}{C_o} \\ &= E \left(1 + \frac{D}{D_o} \sin \omega t\right). \end{aligned}$$

The variation of voltage at the microphone terminals is thus of the same form as the sound wave, if the required conditions are satisfied. Actually of course the inductance cannot entirely

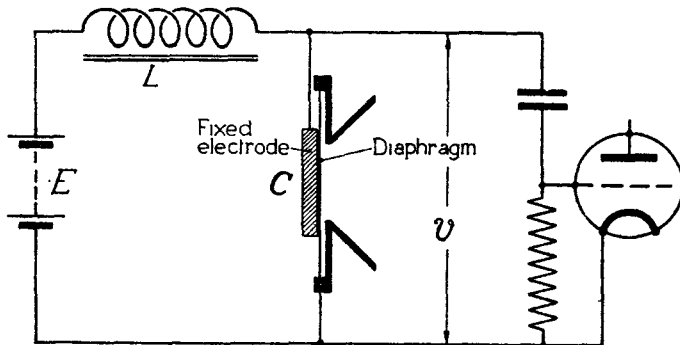


FIG. 3, CHAP. XII.—Condenser microphone.

prevent a variation of charge, and the variation of capacitance will set up a current variation. The result is that v will vary in a more complex manner than the original sound wave, but the additional frequencies introduced will not be of large amplitude, because they are proportional to the second and higher powers of $\frac{D}{D_o}$ and this ratio is always much less than unity.

(ii) In the practical form of condenser microphone the case containing the electrodes is hermetically sealed in order to prevent the entry of moisture and dust, arrangements being made to allow the normal pressure on each side of the diaphragm to be equalized. The microphone possesses a resonant frequency which is usually of the order of from 3 to 5 kc/s, but its response curve is fairly flat, e.g. taking the voltage response at 800 cycles per second as a standard, a typical instrument may give a response of -5 d.b. at 300 cycles per second, and $+10$ d.b. at its resonant frequency, say 4,000 cycles per second. Above this the response falls rapidly, being perhaps -5 d.b. at 8,000 cycles per second. The instrument obviously has a high impedance at speech frequencies and should work into a very high resistance load, such as the input impedance of a carefully designed audio-frequency amplifier. In any event, since the condenser microphone is of very much lower sensitivity than the carbon microphone, at least two extra stages of audio-frequency amplification are required in order to obtain the same output. It is usual to mount

the first stage as near to the microphone itself as possible to avoid the shunting effect of the capacitance of a long twin lead or concentric cable. The comparative insensitivity of the condenser microphone outweighs the advantage of its fairly uniform frequency response, except in circumstances where the space and weight of the ancillary apparatus is of no importance.

The moving coil microphone

12. In this instrument the movement of a light coil of wire in a powerful magnetic field produces an E.M.F. conforming in variation with the sound waves which cause the movement of the coil. The magnetic circuit is so arranged that the only air gap is an annular space two to four centimetres in diameter and about $\cdot 2$ centimetres wide. As shown in fig. 4, the moving coil is wound on a light former, which is carried by a suitable diaphragm in such a manner that the movement of the latter causes a variation of flux through the coil, and the production of a corresponding E.M.F. A fairly uniform response over a wide frequency range is achieved by careful design of the air chamber enclosed by the magnetic system and diaphragm; taking the

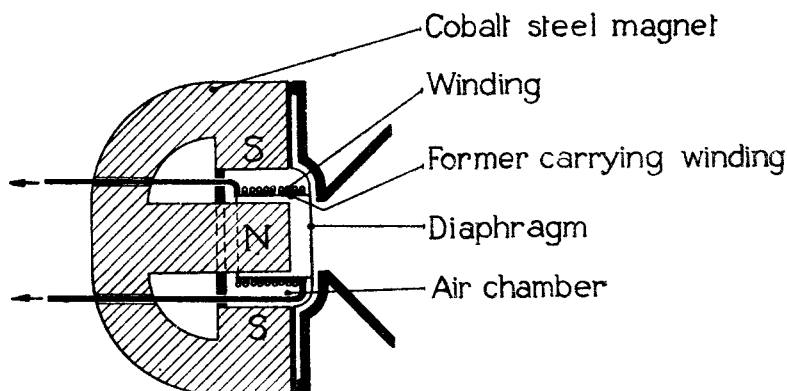


FIG. 4, CHAP. XII.—Moving coil microphone.

response at 800 cycles per second as standard the response of a typical instrument may be -3 d.b. at 50 cycles per second, $+5$ d.b. in the neighbourhood of the resonant frequency (about 3,000 cycles per second) and falling rapidly above 10,000 cycles per second. One or more subsidiary resonances may be observed between 3,000 and 10,000 cycles per second. The diaphragm is actuated partly by the change of pressure and partly by the air particle velocity and consequently the instrument is somewhat directional particularly at the higher frequencies, maximum response being obtained when the sound wave impinges perpendicularly upon the diaphragm. The instrument has a low impedance and may be transformer-coupled to the first valve of a speech amplifier. It is rather more sensitive than the condenser microphone, but not nearly so sensitive as the carbon microphone. Both the condenser and moving coil microphones have the advantage of a much lower noise level compared with the carbon microphone, but unless the additional amplifier stages are carefully designed and operated these advantages may be offset by the additional amplifier noise.

MODULATION

Necessity for carrier wave

13. (i) The microphone was originally intended for use in telephonic communication between places connected by wire, and the audio-frequency variations of current resulting from its operation were conveyed from point to point in their original form. It is not possible, however, to radiate power at audio-frequencies over any appreciable distance, for the energy radiated in the form of electro-magnetic waves by an open oscillator is proportional to the square of the

and a complete cycle of this variation occurs in a certain time, say T_a seconds. The frequency of this variation is $\frac{1}{T_a}$ cycles per second and may be denoted by f_a , which is the audio-frequency at which modulation is taking place. The amplitude at any instant, say t seconds from the instant A at which modulation commences, is $\mathcal{I}_c + \mathcal{I}_a \sin 2\pi f_a t$, or if $\omega_a = 2\pi f_a$, the amplitude at this instant is $\mathcal{I}_c + \mathcal{I}_a \sin \omega_a t$. Although the amplitude of the current is no longer constant, its frequency is still f_c , and the modulated wave form may be completely represented by the equation

$$i_m = (\mathcal{I}_c + \mathcal{I}_a \sin \omega_a t) \sin \omega_c t \quad \dots \quad (1)$$

The ratio $\frac{\mathcal{I}_a}{\mathcal{I}_c}$ is called the depth of modulation, or, if multiplied by 100, the percentage of modulation. In the diagram below (fig. 5) $\frac{\mathcal{I}_a}{\mathcal{I}_c}$ is equal to 0.5 and the oscillation is said to be modulated 50 per cent. It is usual to denote the depth of modulation by the symbol K . Hence equation (1) may be written

$$i_m = (1 + K \sin \omega_a t) \mathcal{I}_c \sin \omega_c t \quad \dots \quad (2)$$

(iii) Instead of regarding the modulated waveform as having a single frequency but varying amplitude, it is convenient to regard it as the sum of a number of radio-frequency oscillations

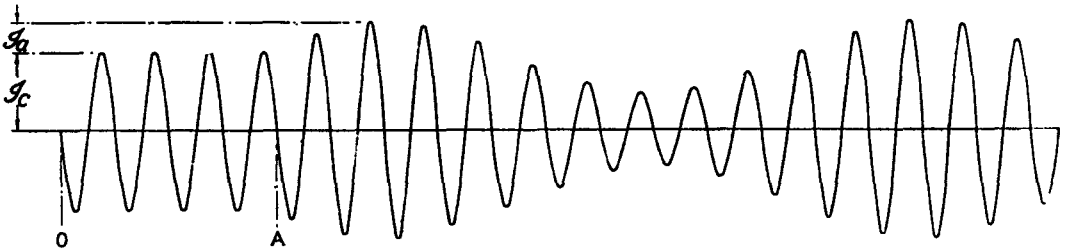


FIG. 5, CHAP. XII.—Amplitude modulated wave.

of various frequencies, each of which is of constant amplitude. There is no subterfuge in such a transformation; it has been stated repeatedly that no matter how complex a wave-form may be, provided only that it is repetitive, it may be proved to be the sum of a number of sinusoidal components. In the present instance, we may make use of an expression developed in Chapter V, namely

$$2 \sin P \sin Q \equiv \cos (P - Q) - \cos (P + Q)$$

The sign $\equiv$ has been used in order to emphasise that this is not an equation but an identity, that is the two members are always equal no matter what values are allotted to P and Q . The equation representing the modulated oscillation, viz. $i_m = (\mathcal{I}_c + \mathcal{I}_a \sin \omega_a t) \sin \omega_c t$ may be written

$$i_m = \mathcal{I}_c \sin \omega_c t + \mathcal{I}_a \sin \omega_c t \sin \omega_a t.$$

By the above identity, therefore,

$$i_m = \mathcal{I}_c \sin \omega_c t + \frac{\mathcal{I}_a}{2} \cos (\omega_c - \omega_a) t - \frac{\mathcal{I}_a}{2} \cos (\omega_c + \omega_a) t. \quad \dots \quad (3)$$

The modulated oscillation is thus shown to be the sum of three component oscillations, each of constant amplitude, namely:—

(a) The original carrier current, of amplitude $\mathcal{I}_c$ and frequency $f_c = \frac{\omega_c}{2\pi}$.

CHAPTER XII.—PARA. 15

(b) An oscillation of constant amplitude $\frac{\mathcal{P}_a}{2}$, the frequency of which is less than that of the carrier by an amount equal to f_a , the frequency of the modulation. This is called the lower side-frequency.

(c) An oscillation of constant amplitude $\frac{\mathcal{P}_a}{2}$, the frequency of which is greater than that of the carrier by an amount equal to f_a . This is called the upper side-frequency.

It will be noted that the modulation frequency f_a is in the audio-frequency range, while f_r is a radio-frequency; the upper and lower side-frequencies being $(f_r + f_a)$ and $(f_r - f_a)$ respectively are both radio-frequencies.

15. In order to prove that the sum of the above three component oscillations gives the modulated oscillation of fig. 5, fig. 6 has been prepared. The two oscillations $\frac{\mathcal{P}_a}{2} \cos (\omega_c + \omega_a) t$ and $\frac{\mathcal{P}_a}{2} \cos (\omega_c - \omega_a) t$ are shown at (a), the former oscillation commencing 180° out of phase with the latter in order to comply with the negative sign prefixed to it in equation (3) above.

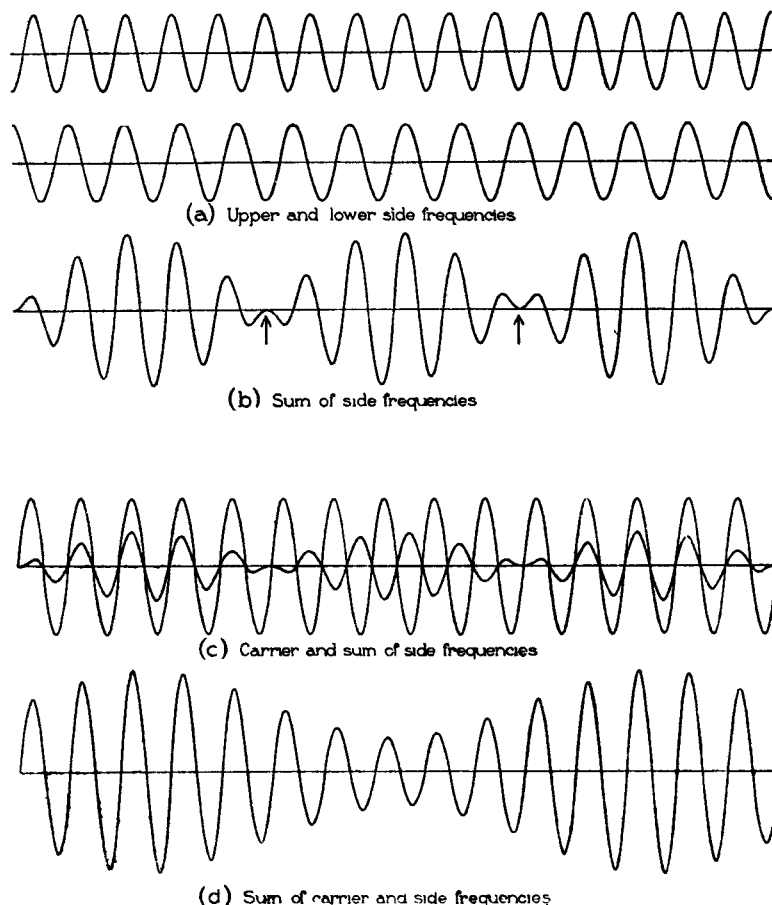


FIG. 6, CHAP. XII.—Synthesis of modulated wave.

The sum of these two is shown at (b). This result is identical with that obtained in heterodyne reception when the incoming and local oscillations are of equal amplitude. The frequency of the individual waves forming each "beat" is the mean of the two component frequencies, namely $\frac{1}{2\pi} \left\{ \frac{(\omega_c + \omega_s) + (\omega_c - \omega_s)}{2} \right\} = \frac{\omega_c}{2\pi} = 2f_c$, i.e. it is equal to that of the carrier. Again, the number of beats per second is equal to the difference between the two frequencies, being equal to $\frac{1}{2\pi} \left\{ (\omega_c + \omega_s) - (\omega_c - \omega_s) \right\} = \frac{2\omega_s}{2\pi} = 2f_s$, hence there are two complete beats in the time occupied by one audio-frequency cycle. A point of great importance in the theory of modulation, although of no significance in heterodyne reception, is the change of phase which occurs at the instant at which one beat is concluded and the next is commenced. These instants are indicated by arrows in the figure.

16. In fig. 6c, the beats formed as described above are superimposed upon the carrier oscillation in order to show the effect of this change of phase. During the time taken by the first beat, the carrier and beat oscillations are in phase, and the amplitude

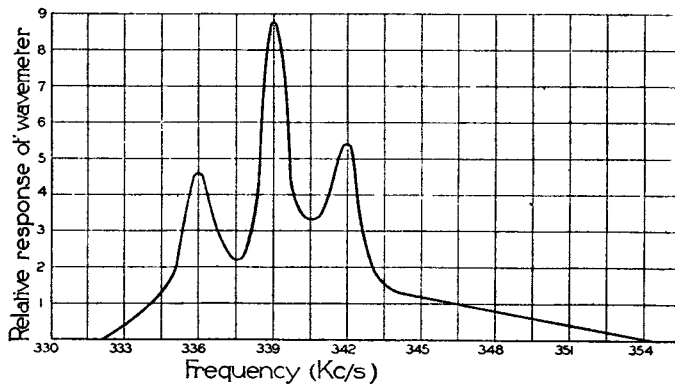


FIG. 7, CHAP. XII.—Response of wavemeter to carrier and side frequencies.

of the resultant, which is initially $\mathcal{I}_c$, gradually rises to a value $\mathcal{I}_c + \mathcal{I}_s$, and then falls again, reaching its original value $\mathcal{I}_c$ at the end of the first beat. Owing to the change of phase at this instant, during the next beat the carrier and beat oscillations are 180° out of phase with each other, and the amplitude of their sum falls to $\mathcal{I}_c - \mathcal{I}_s$ at the middle of the beat. The amplitude then increases, and attains the value $\mathcal{I}_c$ when the end of the beat is reached, the foregoing cycle of events being then repeated. The sum of the three component oscillations is shown in fig. 6d and it is seen to be identical with the original modulated oscillation in fig. 5. When an aerial is carrying an oscillatory current, sinusoidally modulated by a fairly high audio frequency, the carrier oscillation and the upper and lower side-frequencies can be separately identified by a wavemeter, provided that its oscillatory circuit has a high magnification. Fig. 7 shows the response of such a wavemeter in a particular instance, the carrier frequency being 339 kc/s and the modulating audio frequency 3 kc/s. It is seen that, in accordance with the foregoing analysis, the side-frequencies are respectively 3 kc/s above and below the carrier frequency. This experiment was actually performed to demonstrate the objective existence of the side-frequencies. The fact that the wavemeter response at the upper side-frequency is somewhat higher than that at the lower does not mean that the amplitudes of these components are unequal, but merely signifies that the magnification of the wavemeter—and probably that of the aerial circuit also—is slightly greater at 342 kc/s than at 336 kc/s.

CHAPTER XII.—PARAS. 17-18

17 We have now seen that the modulation of a carrier by a single audio frequency results in the production of two radio-frequency oscillations in addition to the original carrier. Suppose now that the carrier is simultaneously modulated by two audio frequencies. Each of these will give rise to an upper side-frequency and a lower side-frequency and, in general, every audio-frequency oscillation by which the carrier is modulated gives rise to a pair of side-frequencies, those higher than the carrier frequency being collectively referred to as the upper side-band, and those lower than the carrier frequency being called the lower side-band. For example, if the carrier frequency is 100 kc/s and it is simultaneously modulated by frequencies of 250, 480 and 1,200 cycles per second, the modulated oscillation would contain all the following components :—

(i) The carrier frequency		1,000,000 cycles per sec.
(ii) The upper side-band		1,000,000 + 250 = 1,000,250 1,000,000 + 480 = 1,000,480 1,000,000 + 1,200 = 1,001,200
(iii) The lower side-band		1,000,000 - 250 = 999,750 1,000,000 - 480 = 999,520 1,000,000 - 1,200 = 998,800

The complex vibrations of speech and music can be resolved into a number of sinusoidal variations, and when a carrier oscillation is modulated in accordance with the frequencies of these sound waves, the side-bands consist of a pair of radio-frequency oscillations for each constituent (sinusoidal) audio-frequency component of the sound wave. When a carrier oscillation is subjected to modulation by several audio-frequencies simultaneously, some care is needed in speaking of the depth or percentage of modulation, in fact these terms can only be used with accuracy when dealing with sinusoidal modulation by a single frequency. The point is best illustrated by actual examples, such as are shown in figs. 8, 9 and 10, to which further reference will be made.

Power in carrier and side-bands

18. The total energy contained in a modulated wave is the sum of the energies carried by the components of different frequencies, so that if, for example, we have an aerial circuit whose radiation resistance is R ohms, and I_0 is the R.M.S. carrier current, the power radiated by the carrier alone is $I_0^2 R$ watts. Now suppose the wave to be modulated to a depth of unity, the R.M.S. value of each side component will be $\frac{I_0}{2}$, and the total power radiated by the side-

frequencies will be $\left(\frac{I_0}{2}\right)^2 R \times 2 = \frac{I_0^2 R}{2}$. The total power radiated in the completely modulated

wave is therefore $(1 + \frac{1}{2}) I_0^2 R = \frac{3}{2} I_0^2 R$, of which only one-third is carried by the side-frequencies and two-thirds by the carrier frequency. Now instead of a completely modulated wave, for which $K = 1$, let us consider one modulated to a depth of less than unity. The carrier will radiate $I_0^2 R$ watts, as before, and each side-frequency, having an R.M.S. value $\frac{K I_0}{2}$, will

radiate $\left(\frac{K I_0}{2}\right)^2 R$ watts, the total side-band power being therefore $\frac{K^2 I_0^2 R}{2}$ watts. The

total power contained in the wave is now equal to $I_0^2 R \left(1 + \frac{K^2}{2}\right)$ so that the power in the side

frequencies is $\frac{K^2}{2 + K^2}$ of the total power. Since the required intelligence is conveyed entirely by the side-frequencies, the effective signalling range is rapidly reduced by a reduction in the depth of modulation. From this point of view it would appear desirable to modulate to a depth of unity, but in practice it is usual to aim at not more than about 80 per cent. This is partly on

account of certain difficulties which arise in reception, partly because provision must be made for a considerable variation in the intensity of the modulating sound, but chiefly because in practice the latter is rarely of sinusoidal wave-form. For distortionless transmission, the envelope of the radio-frequency oscillation must be identical with the wave-form of the audio-frequency variation. This will not be so, even with sinusoidal modulation, if the variation of amplitude exceeds the amplitude of the carrier. Thus in fig. 8a, the variation of amplitude is equal to $1.5 \mathcal{A}_c$; the envelope of the resulting radio-frequency oscillation is far from sinusoidal, and after detection would be found to contain a pronounced second harmonic. In practice however the wave-form shown in fig. 8a cannot be produced by ordinary methods of modulation. Under the conditions indicated, the usual result is a total cessation of oscillation during the "overlap" period, as in fig. 8b. Such a wave is said to be over-modulated.

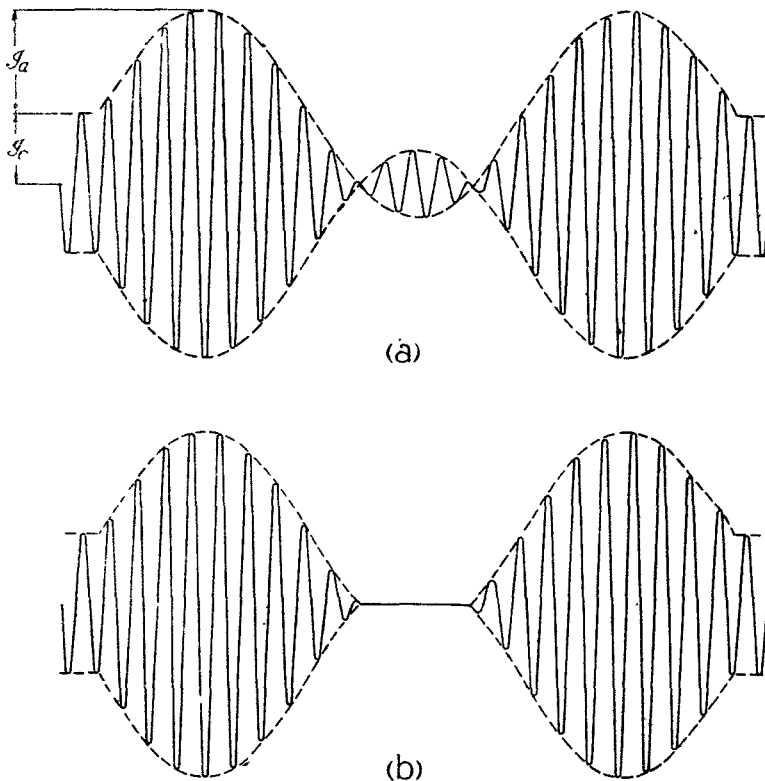


FIG. 8, CHAP. XII.—Over-modulation.

19. Assuming that a wave is completely modulated, more energy is radiated in the side-bands of a flat-topped wave than in one of peaky form. Figs. 9 and 10 show the components and resultant modulated wave when modulation is caused by a fundamental audio frequency and its third harmonic, the relative phase in fig. 9 corresponding to fig. 1a, and in fig. 10 to that of fig. 1b. Taking fig. 9 first, the carrier wave has an amplitude of $\mathcal{A}_c$, say, and the fundamental side frequency $\mathcal{A}_1$ is equal to $\frac{\mathcal{A}_c}{2}$. It can be shown that a third harmonic having an amplitude $\mathcal{A}_3 = \frac{2}{5} \mathcal{A}_1 = \frac{1}{5} \mathcal{A}_c$ can be added without over-modulation. Then if the radiation resistance of

CHAPTER XII.—PARA. 20

the aerial is R as before, the power in the carrier is $\frac{\mathcal{J}_0^2 R}{2}$, in the fundamental side-frequencies $2 \left(\frac{\mathcal{J}_1^2 R}{2 \times 2^2} \right)$, and in the third harmonic side-frequencies $2 \left(\frac{\mathcal{J}_3^2 R}{2 \times 5^2} \right)$. Thus the power in the side-bands is to the power in the carrier as $2 \left(\frac{1}{4} + \frac{1}{25} \right)$ is to 1 or .58 to 1. This may be compared with .33 to 1 for sinusoidal modulation (paragraph 18). If, however, the phase of the third harmonic is reversed, giving rise to the peaky wave form shown in fig. 10, and the third harmonic is still to have an amplitude equal to $\frac{2}{5} \mathcal{J}_1$, the amplitude of the fundamental itself must be less than $\frac{\mathcal{J}_0}{2}$, otherwise over modulation will occur. A simple calculation shows that for complete modulation, $\mathcal{J}_1 = \frac{5}{14} \mathcal{J}_0$ and $\mathcal{J}_3 = \frac{2}{5} \mathcal{J}_1 = \frac{1}{7} \mathcal{J}_0$, so that the power in the side-bands is to the power in the carrier as $2 \left\{ \left(\frac{5}{14} \right)^2 + \left(\frac{1}{7} \right)^2 \right\}$: 1 or .296 to 1. From these two examples it may be deduced that a perfectly square-topped wave would have the greatest ratio of side-band to carrier power; it can be shown that if such a wave is completely modulated the side-bands contain the same amount of power as the carrier.

Increase of aerial current during modulation

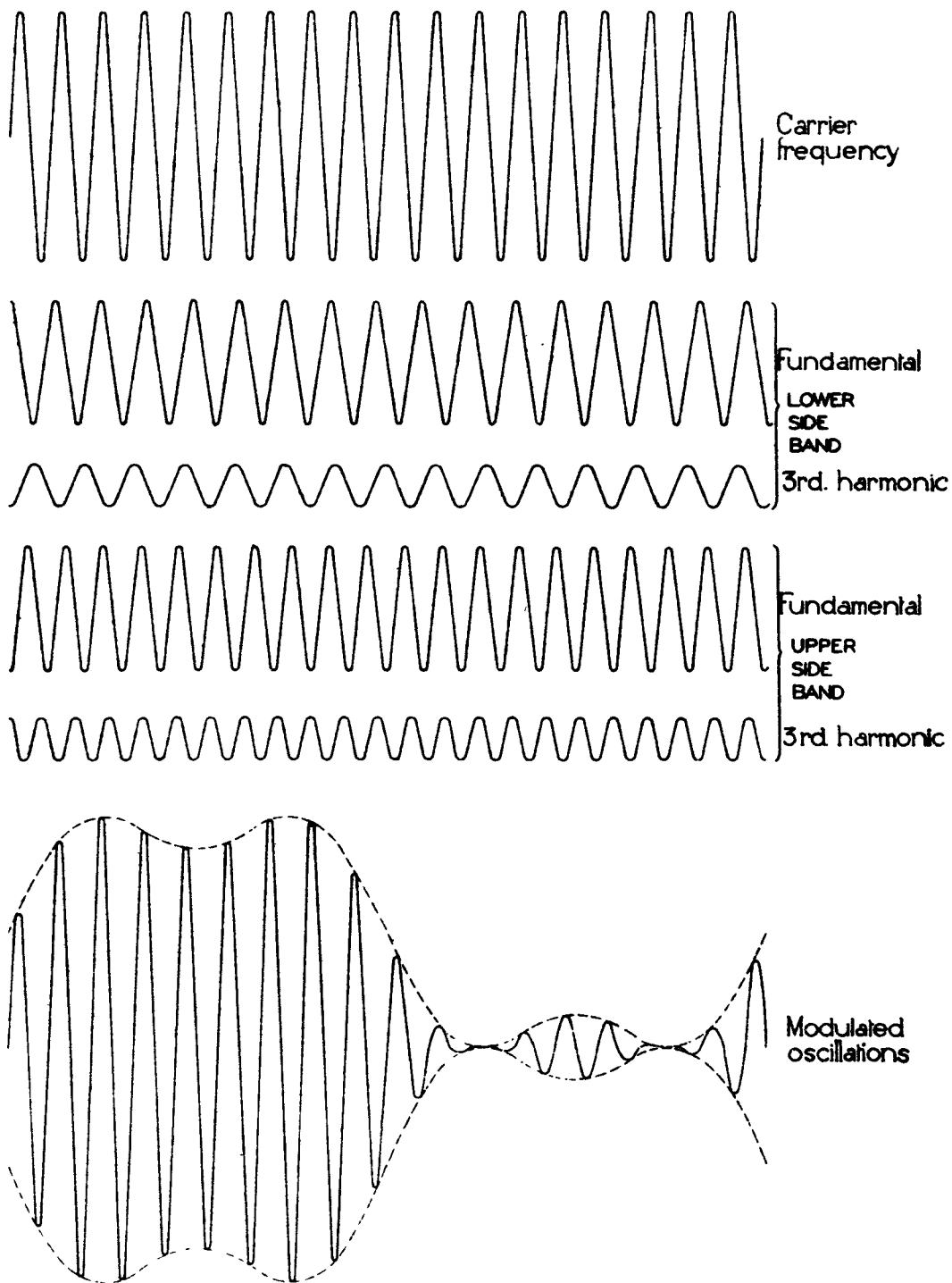
20. Since during modulation the power in the aerial circuit is continually varying, the aerial current must also vary. In paragraph 18 it was shown that the power in the aerial circuit is proportional to $I_0^2 \left(1 + \frac{K^2}{2} \right)$ where I_0 is the unmodulated R.M.S. aerial current. It follows therefore that during modulation the R.M.S. aerial current will rise to some value I_M where

$$I_M = I_0 \sqrt{1 + \frac{K^2}{2}}$$

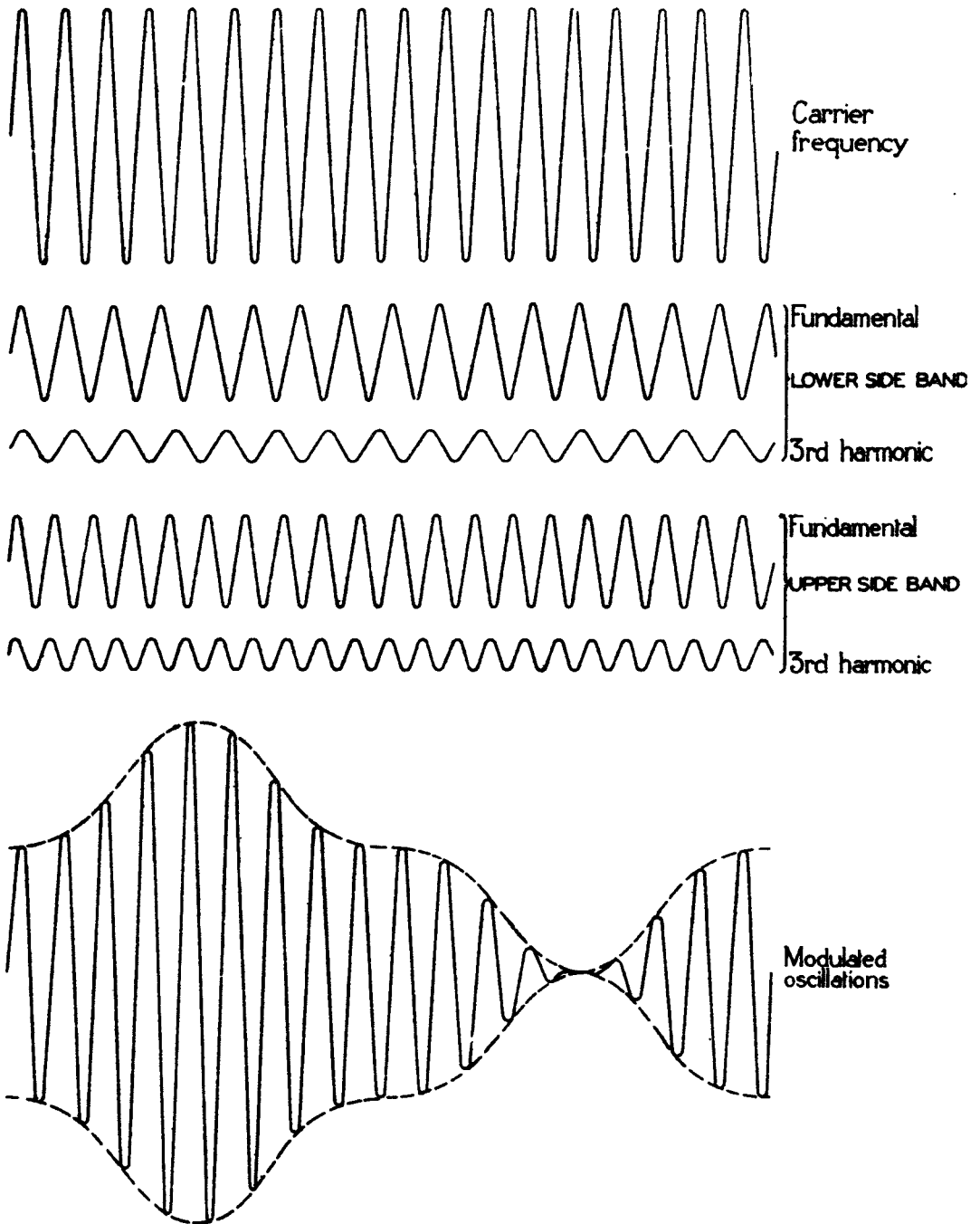
Thus if $K = 1$, the aerial ammeter will indicate an increase of aerial current amounting to only 22.5 per cent; for lower depths of modulation the increase in aerial current will be still less marked. The formula given above is strictly applicable only when the wave is sinusoidally modulated. If the modulation is caused by a flat-topped audio-frequency wave, complete modulation will give rise to a 40 per cent. increase in aerial current. During an actual R/T transmission, however, the aerial current should never show even a momentary increase of this order, for the depth of modulation and audio-frequency wave-form are never constant over periods comparable with the "lag" of the aerial ammeter. We may estimate the probable increase of aerial current, under correct working conditions, as follows. Since it is desirable to limit the peak modulation to 80 per cent., the mean depth of modulation may be taken as 40 per cent. Taking the average depth over a period of several seconds, as distinct from the mean depth, we must allow for intervals between words, say 20 per cent. of the total time, thus the average depth will be about .8 of the mean depth. Inserting $K = .32$ in the formula, we find

$$\begin{aligned} I_M &= I_0 \sqrt{1 + \frac{.32^2}{2}} \\ &= 1.05 I_0. \end{aligned}$$

Although obtained by a very approximate method, this figure agrees with the results obtained on a correctly adjusted transmitter during speech modulation. A 5 per cent. increase of aerial current is not noticeable in a transmitter fitted with the usual thermo or hot-wire ammeter; during modulation, the pointer of such an instrument may be in a state of barely perceptible vibration, but the occurrence of violent movements indicates that the aerial oscillations are



**WAVE MODULATED BY FUNDAMENTAL AND
+ve. THIRD HARMONIC**



**WAVE MODULATED BY FUNDAMENTAL AND
-ve THIRD HARMONIC**

**FIG. 10
CHAP. XII**

intermittent. This in turn shows that serious over-modulation is taking place, as will be seen on reference to fig. 8b. Similar considerations apply to the modulated stages of frequency-controlled transmitters, without regard to the manner in which the modulation is performed.

Methods of modulation

21. Assuming that we have at our disposal a generator of radio-frequency oscillations, such a simple C.W. transmitter in which the key terminals have been bridged by a conducting link, the radiated wave may be modulated by several different methods. In discussing these it will be assumed that a carbon microphone is used to convert the sound waves into electrical impulses in some portion of the circuit.

(i) *Variation of aerial resistance.*—The simplest method of modulation is to connect the microphone directly in series with the aerial or in the earth lead as shown in fig. 11a. When no speech is taking place, a continuous wave is radiated from the aerial, but on speaking into the microphone, the resulting variation in its resistance causes a corresponding

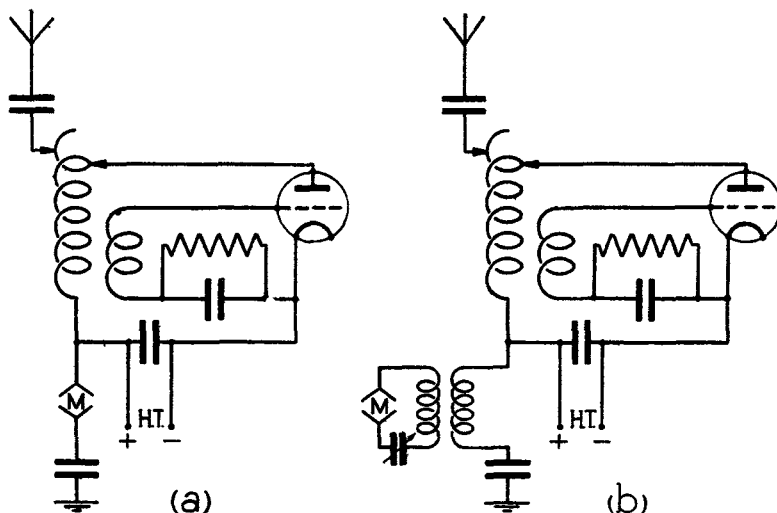


FIG. 11, CHAP. XII.—Modulation by variation of aerial resistance.

variation in aerial current and therefore of the power radiated. The amplitude of the aerial current will vary in accordance with the wave-form of the sound waves impressed upon the microphone and after detection at the receiver, an audio-frequency current of the same wave-form will be obtained. This will of course actuate the reproducing device, i.e. telephones or loud speaker, and so the original speech wave-form will be reproduced. This method is of little practical importance for several reasons. First, if an ordinary carbon microphone is used, its resistance is added to that of the aerial circuit, and this will in itself necessitate a large increase in power input to the transmitter if the same carrier power is required as is obtained in the absence of the microphone. Second, the microphone must be capable of carrying the whole aerial current, and this prohibits the use of a carbon microphone of ordinary design, which will only handle a feed current of a few hundred milliamperes without overheating. Third, the variation of microphone resistance is small compared with its mean resistance, and the variation of aerial current will be small compared with its mean value, hence only a small depth of modulation is obtainable. Some improvement is obtainable by coupling the microphone to the aerial circuit indirectly, (fig. 11b), but even then the utility of the method is limited to low powers and low depths of modulation.

CHAPTER XII.—PARA. 22

(ii) *Variation of anode potential.*—As an alternative, the microphone may be caused to vary the anode supply voltage to the oscillator valve; a simple circuit by which this may be achieved is given in fig. 12, in which a self-oscillatory transmitter provides the carrier frequency aerial current. In series with the H.T. supply is included the secondary winding of an iron-core transformer, the primary circuit of which contains a microphone and battery. When the microphone is quiescent the amplitude of the aerial current depends upon the H.T. supply voltage, and if the latter is constant the carrier wave will be of unvarying amplitude. On speaking into the microphone, the variations in its resistance cause variations of current in the primary winding of the transformer and consequent variation of magnetic flux. A corresponding E.M.F. will be induced in the secondary winding, the wave-form of which will closely resemble that of the original sound wave. The supply voltage to the oscillator valve is no longer constant, but consists of this induced E.M.F. superimposed upon the H.T. supply voltage. As the amplitude of the aerial current is proportional to the anode-filament P.D. it is apparent that the radiated wave will be modulated at the speech frequency. This method is of no practical importance owing to the limited depth of modulation obtainable; as a first approximation it may be

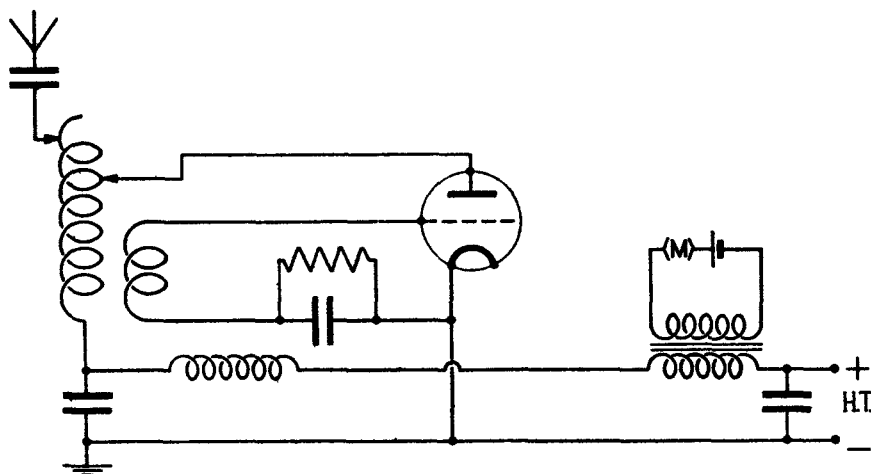


FIG. 12, CHAP. XII.—Modulation by variation of anode voltage.

assumed that in order to attain 100 per cent. modulation the peak value of the transformer secondary E.M.F. should be equal to the constant voltage derived from the H.T. supply. This is impossible with oscillators having a supply voltage higher than about 50 volts, but the method forms the basis of the system known as choke control modulation.

(iii) *Choke control modulation.*—This system is similar in principle to that just described but may be utilized in transmitters of the highest powers used in service or commercial practice.

(iv) *Grid bias modulation.*—This system is suitable for use in low power transmitters, particularly where the primary function is C.W. or I.C.W. telegraphy, and R/T is only occasionally required.

The two latter are in extensive use in service R/T transmitters.

Action of choke control modulator

22. (i) The circuit of a simple form of transmitter using this form of modulation is given in fig. 13, in which O is the oscillator valve and M is called the modulator valve. The iron-core inductance L_s , which is called the speech choke, is in series with the anode of the modulator

valve and also in series with the anode of the oscillator valve. The microphone with its transformer is arranged to supply speech-frequency voltages between grid and filament of the modulator valve, and the action may be outlined as follows. When the microphone is quiescent, the anode-filament P.D. of the oscillator valve is constant, and the oscillator valve maintains an unmodulated aerial current. When sound vibrations impinge upon the microphone diaphragm, the grid-filament P.D. of the modulator valve will vary at the frequency of the sound waves. These changes in grid-filament P.D. result in corresponding variations in the anode current of the modulator valve, and consequently of similar but amplified variations of P.D. between the terminals of the iron-core inductance L_3 . As this choke is also in series with the H.T. supply to the oscillator valve, the anode-filament P.D. of the latter will also vary at speech frequency. The amplitude of the aerial current will then vary in like manner, i.e. will be modulated at the frequency of the original sound wave impressed upon the microphone.

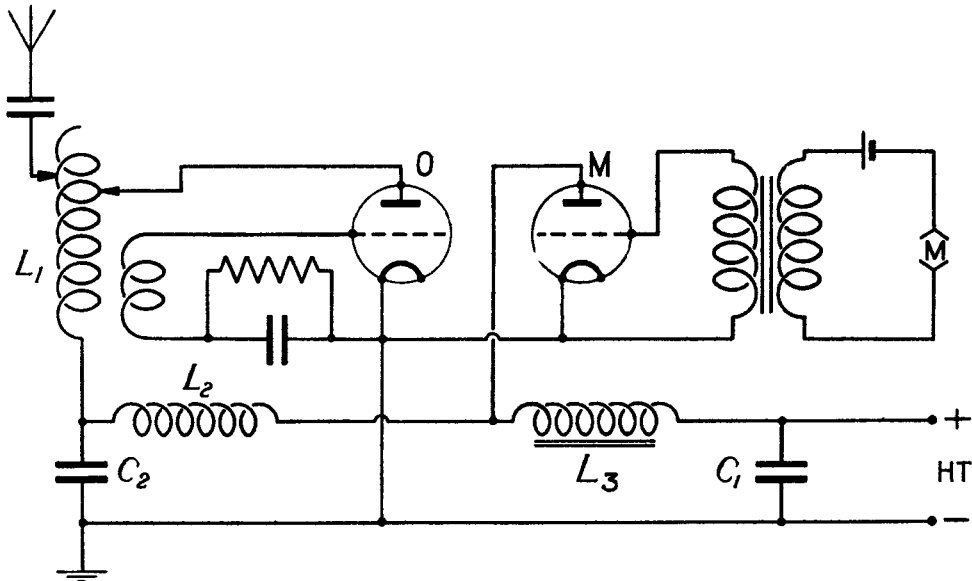


FIG. 13, CHAP. XII.—Choke control modulation.

(ii) Although in this simple explanation of the principle of choke control modulation it is stated that the function of the modulator valve is to set up variations of voltage across the speech choke, in reality the modulator valve is called upon to vary the power supply to the oscillator. It must therefore function as a power amplifier, and the depth of modulation depends upon the amount of power which the modulator valve is able to deliver. It is desirable that the relation between the amplitude of the radio-frequency oscillation and the anode supply voltage to the oscillator valve shall be a linear one, as in fig. 14, in which the amplitude of the oscillatory current in the aerial circuit is plotted against the anode-filament P.D. of the oscillator valve. The "curve" connecting these quantities is a straight line through the origin. If, as in the diagram, the anode-filament P.D. varies sinusoidally at speech frequency, the amplitude of the aerial oscillations will vary in the same manner and will possess a sinusoidal envelope. An approach to this ideal relationship is achieved by operating the oscillator with a large negative grid bias obtained by the condenser and leak method. Under these conditions the amplitude of the oscillatory anode-filament P.D. is only slightly less than the anode supply voltage, so that the desired linear relationship is closely approached.

CHAPTER XII.—PARAS. 23-24

Load on modulator

23. The load into which the modulator valve must deliver power is best appreciated by re-arranging fig. 13. So far as audio-frequency currents are concerned the inductances L_1 and L_2 offer no appreciable impedance, and may be neglected. The output circuit of the modulator then consists of the speech choke L_s , in parallel with which is a capacitance consisting of the mains condenser C_1 and earth condenser C_2 in series. The anode A.C. resistance of the oscillator valve is, in effect, in parallel with the condenser C_2 and acts as a purely resistive impedance at audio-frequencies, hence the equivalent circuit becomes that shown in fig. 15a. As the mains condenser C_1 is always very large compared to the earth condenser C_2 , the circuit may be further simplified, becoming that of fig. 15b. The load impedance into which the modulator valve is called upon to supply power is therefore a very flatly-tuned parallel resonant or rejector circuit, the flat tuning being of course due to the damping imposed by the anode A.C. resistance, r_a , of the oscillator valve.

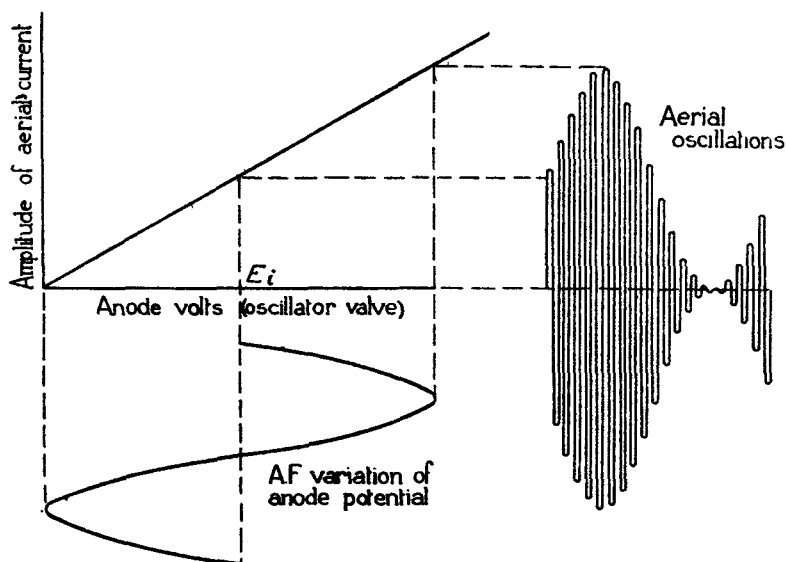


FIG. 14, CHAP. XII.—Ideal relation between anode voltage and aerial oscillations.

Power in carrier and sidebands

24. The power relations for operation under the ideal conditions shown in fig. 14 may be summarized in the statement that the power required to generate the carrier wave is delivered directly to the oscillator valve by the source of H.T. supply, while the power required to generate the side-bands of the modulated wave (although derived from the source of H.T. supply) is delivered in the form of an audio-frequency power output by the modulator valve. When no alternating E.M.F. is generated in the speech choke the power supplied to the oscillator valve is $P_1 = E_i I_1$ where E_i is the voltage of the H.T. supply and I_1 the average anode current. If the oscillator is working at an efficiency of η_o the output power will be $P_o = \eta_o P_1$. During modulation the modulator valve must supply to the oscillator valve the power which is required to vary the anode voltage of the latter; for 100 per cent. sinusoidal modulation the peak value of the voltage induced in the speech choke must be equal to the steady component of the anode voltage E_i . For this depth of modulation the carrier power is twice the power carried by the side-bands and consequently the power output of the modulator valve must be equal to one-half the input power to the oscillator under normal (i.e. unmodulated) conditions. With a lower depth of modulation ($K < 1$) the power output of the modulator will be proportional to K^2 .

Thus if the oscillator takes 60 watts from the source of supply during quiescent periods, the modulator will give 100 per cent. modulation if its output power is 30 watts. If, however, the modulator valve has a power output of only 7·5 watts, it will modulate the carrier to a depth of only $\sqrt{\frac{7.5}{30}} = \cdot 5$ or 50 per cent.

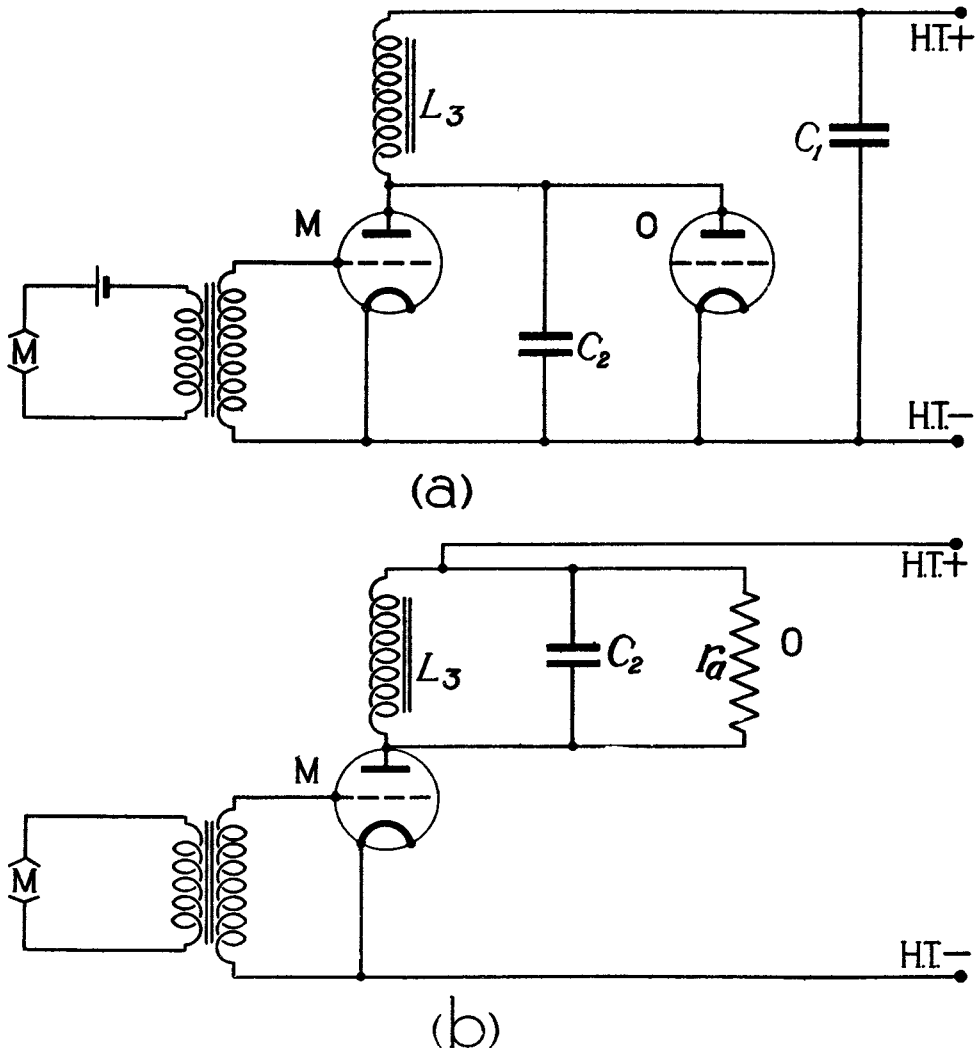


FIG. 15, CHAP. XII.—Equivalent circuit of choke control modulator.

25. If the output is to be completely modulated, the oscillator valve must be operated at an anode voltage E_1 somewhat below that permissible for the generation of undamped oscillations, and the permissible power dissipation is also somewhat smaller. The reduction of anode voltage is necessitated by the fact that during complete modulation the input power is 50 per cent. greater than during quiescent periods, and the anode dissipation increases in the same proportion. Let

CHAPTER XII.—PARA. 25

the permissible dissipation of the oscillator valve be P_L watts, and its input and output powers P_i and P_o watts respectively, then

$$P_o = \eta_o P_i$$

$$P_L = P_i - P_o$$

$$\therefore P_i = \frac{1}{1 - \eta_o} P_L$$

and

$$P_o = \frac{\eta_o}{1 - \eta_o} P_L$$

If this is the power output of the oscillator valve when completely modulated, the power P_c in the carrier will be two-thirds of this and therefore

$$P_c = \frac{2\eta_o}{3(1 - \eta_o)} P_L$$

and the input power to the oscillator is

$$P_i = \frac{P_c}{\eta_o} = \frac{2}{3(1 - \eta_o)} P_L$$

For complete sinusoidal modulation, the output P_m of the modulator valve is one-half this or

$$P_m = \frac{1}{3(1 - \eta_o)} P_L$$

and if η_m is the efficiency of the modulator valve the input to the latter is $\frac{1}{\eta_m}$ times its output. The

input to the modulator must therefore be $P_M = \frac{P_m}{\eta_m}$, or

$$P_M = \frac{1}{3\eta_m(1 - \eta_o)} P_L$$

It is important to remember that if a power amplifier is to operate without appreciable distortion its theoretical efficiency cannot exceed 25 per cent., while in practice it is more likely to be only about 20 per cent. If modulator and oscillator valves are of the same type, it is necessary to use at least three valves in parallel in the modulator valve in order to modulate to a depth of unity the carrier generated by a single valve.

Example :—Using valves having a permissible dissipation P_L of 100 watts, if the oscillator efficiency is 60 per cent., and the modulator efficiency 20 per cent., find the maximum power which can be generated in the carrier, and the power input and output of the modulator in order to modulate the carrier sinusoidally to a depth of unity.

$$\begin{aligned} \text{Carrier power} &= \frac{2}{3} \times \frac{.6}{1 - .6} \times 100 \\ &= \frac{2}{3} \times \frac{6}{4} \times 100 \\ &= 100 \text{ watts.} \end{aligned}$$

$$\begin{aligned} \text{Modulator output} &= \frac{1}{3} \times \frac{1}{1 - .6} \times 100 \\ &= 83.3 \text{ watts.} \end{aligned}$$

The modulator efficiency being 20 per cent., the modulator input will be five times this or 416.5 watts. During periods in which modulation is not actually occurring, the modulator valves are required to dissipate the whole of the input, and therefore, even if four modulator valves are connected in parallel, there will be a slight tendency to overheating.

Variation of load impedance with frequency

26. It has been stated, with reference to figs. 13 and 15, that the anode load impedance, into which the modulator delivers power, is a flatly tuned rejector circuit, and its numerical value and power factor will vary with the frequency. The variation in numerical value, in a particular instance, is shown by the full line curve of fig. 16. This curve was calculated by an approximate graphical method, assuming that $L_3 = 10$ henries, $C_2 = 0.007 \mu F$, r_a (oscillator valve) = 20,000 ohms. At first glance it might appear that the power output of the modulator valve would also vary considerably with the frequency, but this is not so, because the effect of the power factor is not apparent from this curve. The maximum power output will be obtained when the anode A.C. resistance r_a of the oscillator valve is equal to that of the modulator valve, which will be denoted by r_m . Maximum undistorted output, however, is obtained when $r_a = 2r_m$ (see Chapter XI). No matter what the ratio of r_a to r_m may be, the greatest output is always obtained at the resonant frequency of the circuit $L_3 C_2$, and the output falls off rapidly at frequencies below resonance owing to the shunting effect of the speech choke upon the load resistance r_a . At frequencies above resonance, however, the choke has little effect, and the shunting effect of the condenser C_2 causes a slight reduction in output which, however, is not serious except at the very highest audio-frequencies. The curve shown in dotted line in fig. 16 has been calculated for the

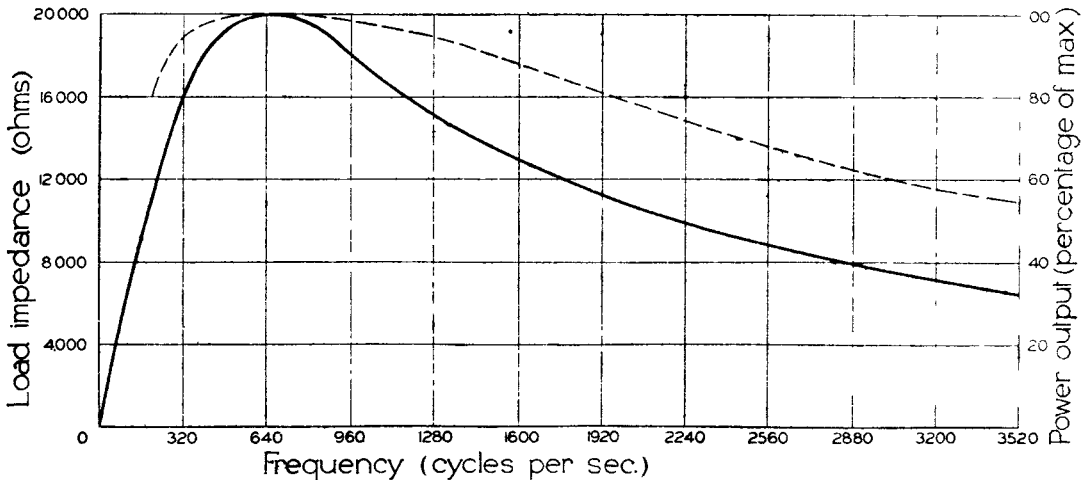


FIG. 16, CHAP. XII.—Load impedance and output power of modulator at various frequencies.

constants already used, and assuming $r_a = 2r_m$. It shows the power output at various frequencies as a percentage of the output at the resonant frequency. An increase in the ratio $\frac{r_a}{r_m}$ will give more uniform frequency response at the expense of a reduction of the peak output, while a decrease in this ratio will give a greater peak output with a corresponding increase in both frequency and amplitude distortion. It should be appreciated that the departure from even response at different frequencies is a measure of the frequency distortion of the circuit. The increase of amplitude distortion which follows a reduction in the ratio $\frac{r_a}{r_m}$ is caused by the increased curvature of the dynamic characteristic of the modulator valve owing to the reduction of load resistance. When the respective values of r_m and r_a are such that the desired matching conditions are not satisfied, either a two-coil audio-frequency transformer, or a suitable auto-transformer may be employed as a coupling between the modulator valve and its output circuit, as shown in figs. 17a and 17b respectively.

CHAPTER XII.—PARA. 27

Modulation of frequency-controlled transmitter

27. With the present-day requirements of constant carrier frequency, the simple choke control modulator circuit with direct aerial excitation has been largely supplanted by circuits in which some form of master-oscillator is employed, the aerial receiving its excitation through one or more stages of radio-frequency power amplification. The modulation may be introduced either in the master-oscillator itself or in one of the amplifier stages. Fig. 18 gives the skeleton diagram of a transmitter in which the master-oscillator stage is modulated. Oscillations of the desired carrier frequency are maintained by the triode T_1 . The speech-frequency voltage from the microphone transformer is first amplified by the sub-modulator valve T_2 which is resistance-capacitance coupled to the modulator valve T_3 . The speech choke L_3 is common to the anode circuits of both the master-oscillator and the modulator valves, thus setting up an audio-

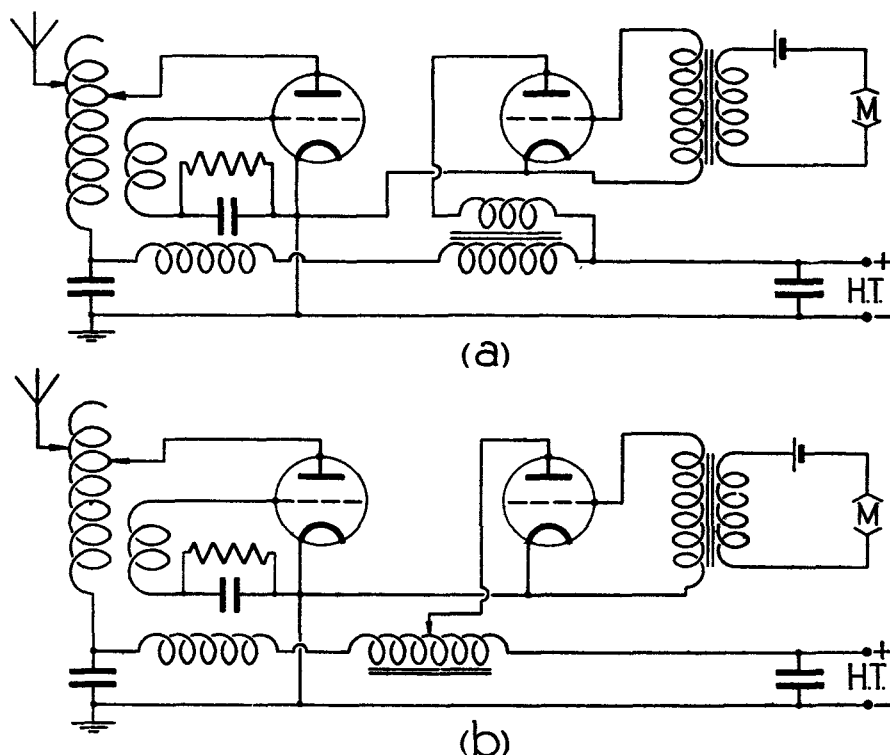


FIG. 17, CHAP. XII.—Alternative circuit arrangements of choke control modulator.

frequency power variation of the radio-frequency oscillation generated in the circuit $L_1 C_1$. The resistance R is fitted in order that the mean operating anode voltage of the master-oscillator shall be below that of the modulator valve. We have already seen that the audio-frequency voltage generated in the speech choke may be nearly but not quite equal to the applied H.T. voltage of the modulator valve, in practice usually about 85 per cent. If then the anode voltage of the master-oscillator is only 85 per cent. of that of the modulator valve, the speech voltage across L_3 will cause the radio-frequency oscillation in $L_1 C_1$ to be modulated to a depth of unity. The master-oscillator derives its grid bias from a condenser and leak resistance. As stated in Chapter IX, this method of biasing an oscillator valve tends to give a constant input conductance and is therefore conducive to frequency stability. In the power amplifier, however, the mean grid bias must be maintained at a constant value irrespective of the instantaneous input grid

swing, otherwise the valve will be overheated during periods of low depth of modulation, and therefore, in this particular circuit, battery bias is used. The frequency stability of a well-designed circuit of this type is considerably better than when direct aerial excitation is used, but where space, weight and cost are not of primary importance, it is capable of further improvement.

28. For the highest degree of frequency stability, it is desirable to maintain the anode voltage of the master-oscillator at a very steady value, modulating at some later stage. Even then it is possible that the variation in input impedance of the modulator valve may react on the master-oscillator in such a manner as to cause frequency variation. This possibility may be removed by the interposition of a buffer or isolator stage between the modulated stage and the master-oscillator, and the circuit diagram of an R/T transmitter of this kind is shown in fig. 19. It comprises a master-oscillator stage, driven by the valve T_1 , and supplying grid excitation to the buffer valve T_5 . The speech voltages are amplified by the submodulator valve T_3 and are applied to the modulator valve T_2 as in the circuit previously discussed. The buffer valve in turn supplies excitation to the modulated amplifier valve T_6 , the speech choke being common to the anode circuits of the valves T_3 and T_6 . The modulated amplifier T_6 is operated under Class C conditions, although if provision is made for complete modulation by a flat-topped audio-

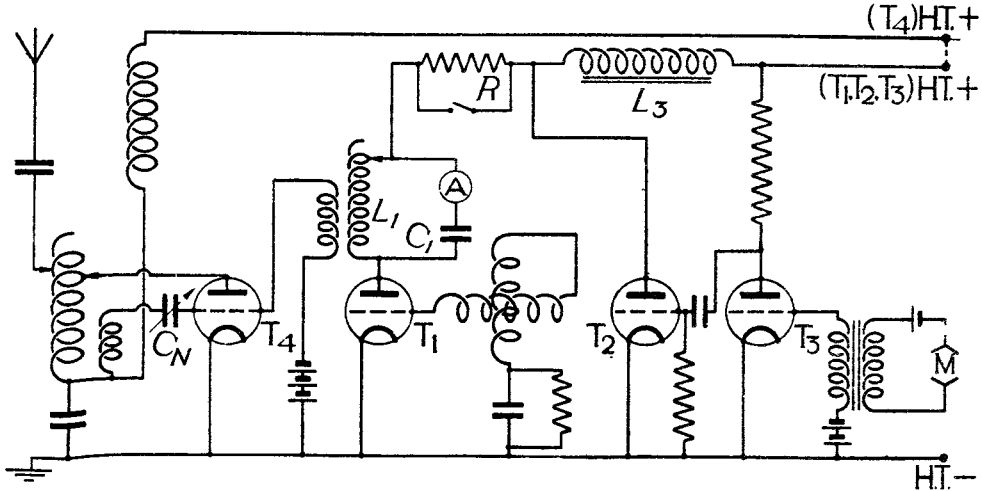


FIG. 18, CHAP. XII.—R/T transmitter with choke control of master oscillator.

frequency wave the average efficiency of this stage will be less than 40 per cent. The modulator must be capable of supplying an amount of power equal to the output of the modulated amplifier, and also of dissipating the whole of its direct current input during periods when no modulation is taking place. Hence the valve T_2 must be capable of dissipating considerably more power than any of the valves immediately associated with it. The modulated output of the valve T_3 is raised to the desired power level by linear amplifiers, which are operated under B class conditions but without running into the grid current region. Two such stages are shown in the diagram. The efficiency of each stage is comparatively low, being only about one-half of that obtainable in a C.W. amplifier. With the valves rated as shown in the diagram, the output of the final linear amplifier would be only about 100 watts. Higher efficiency can be obtained only by operating under conditions which give rise to appreciable distortion.

Grid bias modulation

29. This method is not suitable for high power transmitters but possesses an advantage over choke control modulation in that no separate modulator valve is required. It is possible

CHAPTER XII.—PARA. 30

to employ grid bias modulation in a self-oscillatory transmitter, provided that the mean grid bias is maintained at a constant value, but this application is of little practical importance. The action will be explained with reference to a transmitter controlled by a valve master-oscillator, the essential features of the circuit being shown in fig. 20. In this diagram T_1 is the master-oscillator valve and T_2 the power amplifier. Oscillations are generated in the circuit $L_1 C_1$ and radio-frequency variations of grid-filament P.D. are applied to the amplifying valve via the adjustable coupling between L_1 and L_2 . These voltage variations cause variations of anode current in the amplifying valve, and consequent impulses of voltage in the aerial circuit, the latter being maintained in oscillation by energy drawn from the H.T. supply. The grid-filament potential of the amplifier valve is also varied at speech frequency by means of the microphone, the latter being included in the primary circuit of a suitable step-up transformer. When speech is not taking place, a steady current is established in the circuit comprised by the microphone battery, the microphone itself and the primary winding of the transformer,

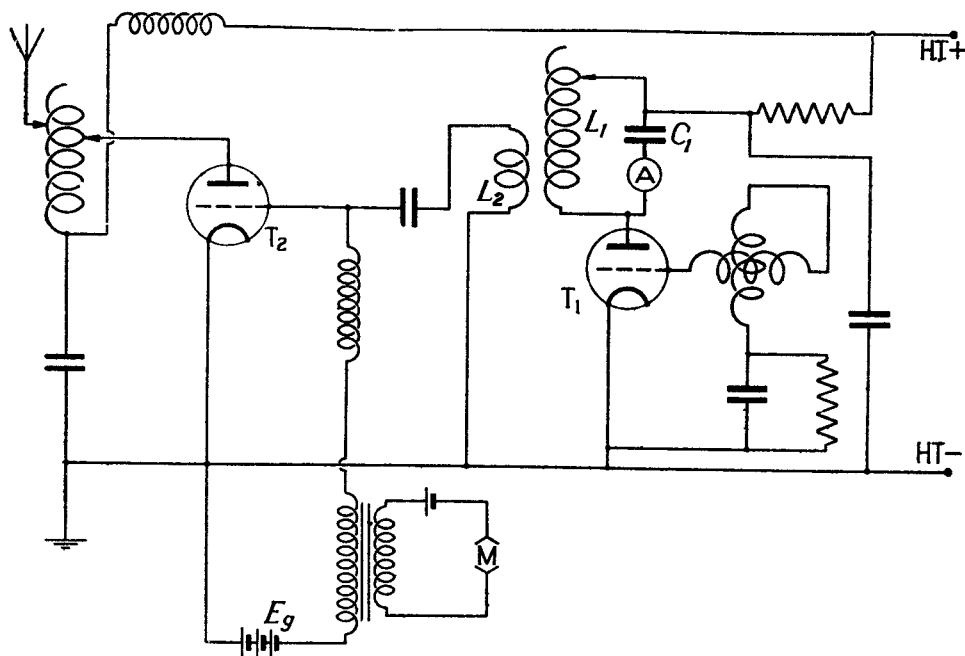
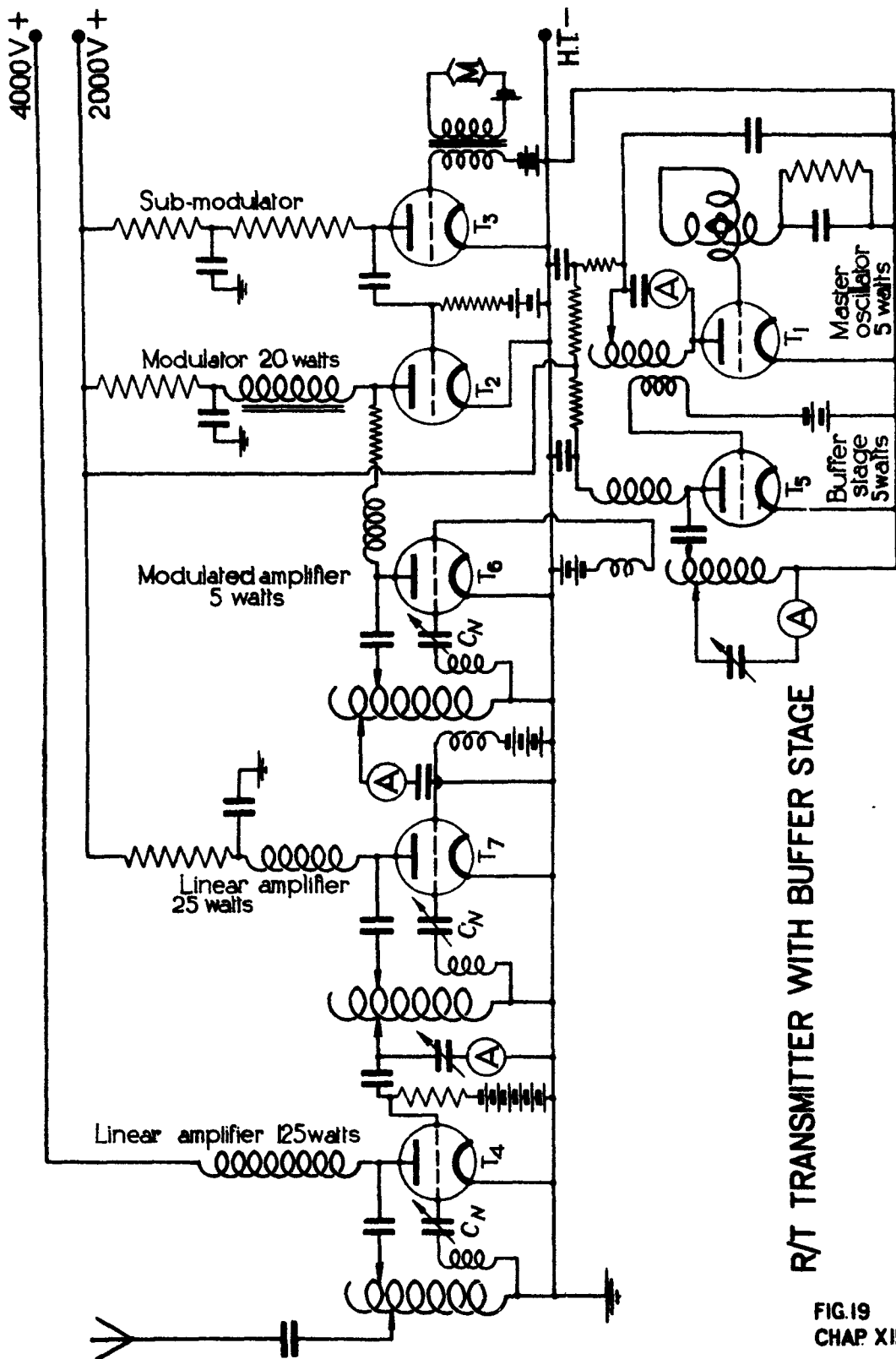


FIG. 20, CHAP. XII.—R/T transmitter with grid bias modulation.

but of course no secondary E.M.F. is generated by this steady current. On speaking into the microphone, however, the variation of resistance causes corresponding changes in the current, which in turn set up a varying flux in the core of the transformer, and an E.M.F. is generated in the secondary winding, the wave form of the secondary E.M.F. closely resembling that of the speech applied to the microphone. As the secondary winding is connected between grid and filament of the amplifier valve, the grid-filament P.D. of the latter is varied at the speech frequency as well as at radio frequency.

30. The effect of the speech-frequency grid-filament P.D. upon the anode current of the valve depends upon (i) the curvature of the $I_a - V_g$ characteristic and (ii) the chosen mean operating point on the curve. Suppose the operating conditions to be as shown in fig. 21, in which the $I_a - V_g$ curve is assumed to be perfectly straight, and the excursions of grid voltage to be limited. The grid-filament voltage consists of three components, namely (i) the steady



R/T TRANSMITTER WITH BUFFER STAGE

bias voltage E_g , (ii) the audio-frequency speech voltage of amplitude $\mathcal{V}_a$, (iii) the radio-frequency voltage of amplitude $\mathcal{V}_r$ due to the master-oscillator. Over the limited range of gr'd voltage shown, the characteristics may be represented by the equation $i_a = g_m v_g$.

As

$$v_g = E_g + \mathcal{V}_a \sin \omega_a t + \mathcal{V}_r \sin \omega_r t$$

$$i_a = g_m [E_g + \mathcal{V}_a \sin \omega_a t + \mathcal{V}_r \sin \omega_r t].$$

Thus the anode current is merely the sum of a steady component, an audio-frequency component and a radio-frequency component and is not of modulated wave-form. We may therefore conclude that if the excursions of grid voltage and anode current are confined to the straight portion of the characteristic, modulation will not be effected.

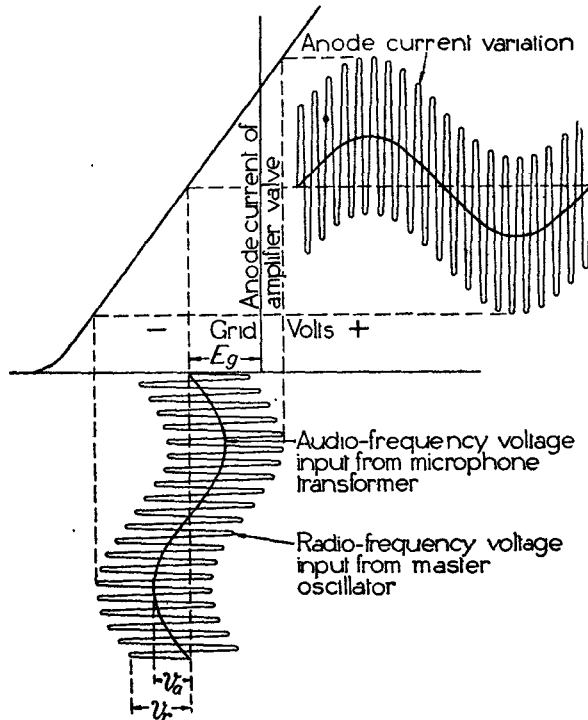


FIG. 21, CHAP. XII.—Operating conditions failing to produce modulation.

“Square-law” modulation

31. The $I_a - V_g$ characteristics of most triodes are approximately parabolic over that portion of the grid voltage lying in the region of negative grid-filament voltage. A fairly complete study of the modulation obtainable with a characteristic of this type is of great assistance in understanding the general principles both of modulation and the detection of modulated waves. The characteristic shown in fig. 22 is a close approximation to the static $I_a - V_g$ curve of a V.T. 25 valve and may be represented by the equation

$$i_a = 360 + 6v_g + \frac{v_g^2}{40} \dots \dots \dots (1)$$

During modulation, the total voltage applied to the grid is

$$v_g = E_g + \mathcal{V}_a \sin \omega_a t + \mathcal{V}_r \sin \omega_r t$$

CHAPTER XII.—PARA. 31

as before ; in the diagram, $E_g = -50$ volts, $\mathcal{V}_s = 25$ volts and $\mathcal{V}_r = 20$ volts. Equation (1) may be expressed in a more general form as

$$i_a = I_o \left(1 - \frac{2}{E_o} v_g + \frac{1}{E_o^2} v_g^2 \right) \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \quad (2)$$

in which I_0 is the anode current at $V_g = 0$ (360 milliamperes in the given curve) and E_0 is the "cut-off" grid voltage, i.e. the grid voltage necessary to reduce the anode current to zero. E_0 must be allotted its correct sign in numerical work. Labour is also economized by writing

$b = -\frac{2}{E_0}$ and $c = \frac{1}{E_0}$, so that equation (2) becomes

$$i_a = I_0 (1 + b v_R + c v_R^2) \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \quad (3)$$

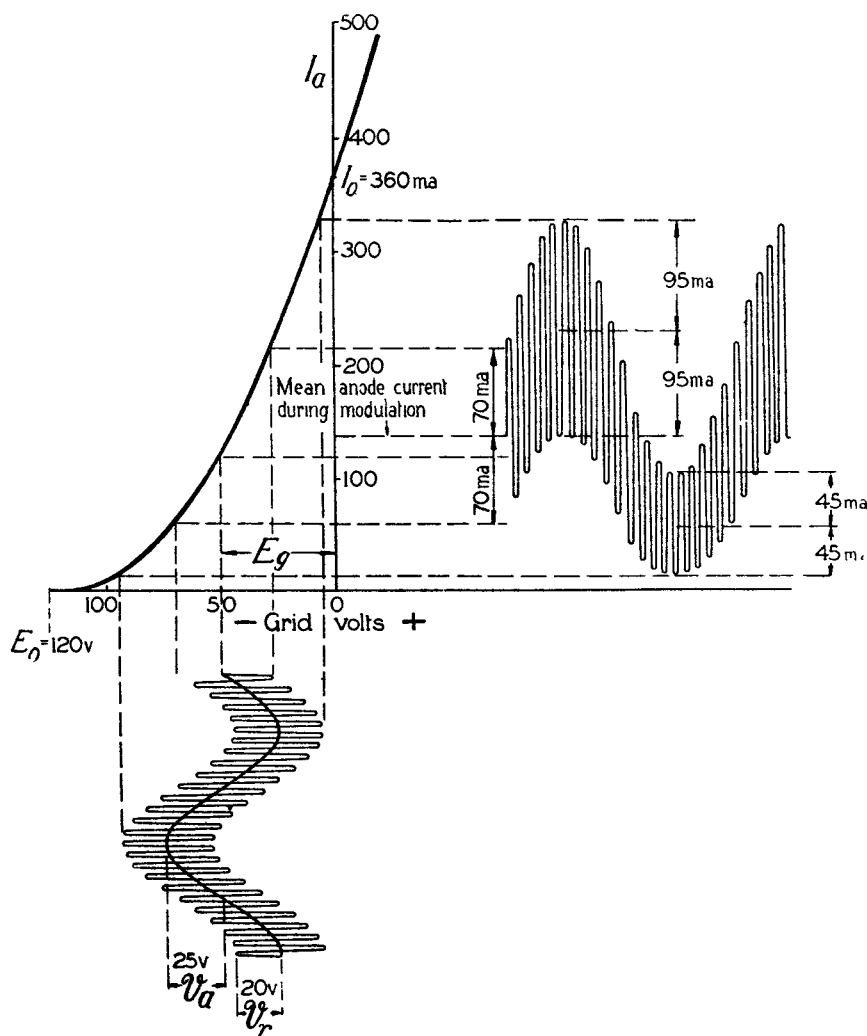


FIG. 22. CHAP. XII.—"Square law" grid bias modulation.

Inserting the above value for v_g , equation (3) becomes

$$\begin{aligned} i_a &= I_o \left[1 + b \left\{ E_g + \mathcal{V}_a \sin \omega_a t + \mathcal{V}_r \sin \omega_r t \right\} \right. \\ &\quad \left. + c \left\{ E_g + \mathcal{V}_a \sin \omega_a t + \mathcal{V}_r \sin \omega_r t \right\}^2 \right] \\ &= I_o \left[1 + b E_g + b \mathcal{V}_a \sin \omega_a t + b \mathcal{V}_r \sin \omega_r t \right. \\ &\quad \left. + c E_g^2 + 2c E_g (\mathcal{V}_a \sin \omega_a t + \mathcal{V}_r \sin \omega_r t) \right. \\ &\quad \left. + c (\mathcal{V}_a \sin \omega_a t + \mathcal{V}_r \sin \omega_r t)^2 \right] \\ &= I_o \left[1 + b E_g + b \mathcal{V}_a \sin \omega_a t + b \mathcal{V}_r \sin \omega_r t \right. \\ &\quad \left. + c E_g^2 + 2c E_g \mathcal{V}_a \sin \omega_a t + 2c E_g \mathcal{V}_r \sin \omega_r t \right. \\ &\quad \left. + c \mathcal{V}_a^2 \sin^2 \omega_a t + 2c \mathcal{V}_a \mathcal{V}_r \sin \omega_a t \sin \omega_r t \right. \\ &\quad \left. + c \mathcal{V}_r^2 \sin^2 \omega_r t \right]. \end{aligned}$$

32. This somewhat complicated expression gives the instantaneous value of the anode current and may now be resolved into a steady component and varying components. Only those containing " $\sin \omega_r t$ " can possibly cause radio-frequency impulses in the anode circuit. Ignoring the common factor I_o , the terms of interest are therefore $b \mathcal{V}_r \sin \omega_r t$, $2c E_g \mathcal{V}_r \sin \omega_r t$, $2c \mathcal{V}_a \mathcal{V}_r \sin \omega_a t \sin \omega_r t$ and $c \mathcal{V}_r^2 \sin^2 \omega_r t$. The latter term is equal to $\frac{c \mathcal{V}_r^2}{2} (1 - \cos 2 \omega_r t)$.

It is responsible for a second harmonic variation of anode current but will not affect the aerial circuit appreciably. Hence the important terms are $(b + 2c E_g) \mathcal{V}_r \sin \omega_r t$ and $2c \mathcal{V}_a \mathcal{V}_r \sin \omega_a t \sin \omega_r t$. The radio-frequency variation of anode current is in fact represented by the equation

$$i_r = I_o [(b + 2c E_g) + 2c \mathcal{V}_a \sin \omega_a t] \mathcal{V}_r \sin \omega_r t$$

$$\text{or} \quad i_r = (1 + K \sin \omega_a t) (b + 2c E_g) I_o \mathcal{V}_r \sin \omega_r t, \quad \dots \dots \dots (4)$$

which will be remembered as the equation of a modulated wave, the carrier amplitude being $(b + 2c E_g) I_o \mathcal{V}_r$, and the depth of modulation K , where

$$\begin{aligned} K &= \frac{2c \mathcal{V}_a}{b + 2c E_g} \\ &= \frac{\frac{2}{E_o^2} \mathcal{V}_a}{-\frac{2}{E_o} + \frac{2 E_g}{E_o^2}} \\ &= \frac{\mathcal{V}_a}{-E_o + E_g} \end{aligned}$$

Bearing in mind that both E_o and E_g are negative and must be allotted their correct signs in numerical calculation this may be written

$$K = \frac{\mathcal{V}_a}{|E_o| - |E_g|}$$

the upright lines denoting that the numerical value of the enclosed quantity is to be inserted without regard to its sign.

In fig. 22 $|E_o| = 120$ volts, $|E_g| = 50$ volts, $\mathcal{V}_a = 25$ volts, hence the depth of modulation is $\frac{25}{120 - 50} = \frac{25}{70} = .3575$ or 35.75 per cent.

33. It must be remembered that the parabolic equation $i_a = I_o (1 + b v_g + c v_g^2)$ only holds if the grid bias and grid swing are so adjusted that the total grid voltage never reaches a negative value exceeding the "cut-off" voltage. The total grid swing is $2(\mathcal{V}_a + \mathcal{V}_r)$. The grid potential base line available to accommodate the negative half of this swing is the difference

CHAPTER XII.—PARA. 34

between E_o and E_g , and so the maximum permissible grid swing is such that $\mathcal{V}_o + \mathcal{V}_r = |E_o| - |E_g|$. With the mean bias fixed at -50 volts, as in the diagram (fig 22) $|E_o| - |E_g| = 70$, and $\mathcal{V}_a + \mathcal{V}_r$ may be 70 volts but not more. Provided that operation takes place wholly within the parabolic region, therefore, the depth of modulation with the maximum permissible grid swing is

$$K = \frac{\mathcal{V}_a}{\mathcal{V}_a + \mathcal{V}_r}$$

which approaches unity as the amplitude of the radio-frequency component of input voltage approaches zero. The amplitudes of the radio-frequency components of anode current will then of course also approach zero. For maximum modulated output we require that the modulation term, $2c\mathcal{V}_a\mathcal{V}_r$, shall be as large as possible consistent with the maximum permissible value of $\mathcal{V}_a + \mathcal{V}_r$. It is apparent that the maximum modulated output will occur when $\mathcal{V}_a$ and $\mathcal{V}_r$ are equal, and the depth of modulation is then only 50 per cent. The amplitude of the carrier-frequency component of anode current has been shown to be $(b + 2cE_g)\mathcal{V}_r I_o$. In

the given example this is $\left(\frac{1}{60} - \frac{2 \times 50}{14,400}\right) \times 20 \times 360 = 70$ milliamperes, which agrees with the

value shown in the diagram. The amplitude of each side-band component is $\frac{2c\mathcal{V}_a\mathcal{V}_r}{2}$ or 12.5 milliamperes, which however is not immediately obvious from an examination of the diagram because the large amplitude of the audio-frequency variation of anode current tends to mask the amplitude of radio-frequency variation. The value can be found as follows. The total variation of anode current, at the positive peak of the audio-frequency cycle, is 190 milliamperes, and at the negative peak is 90 milliamperes. The amplitude of each side-band is

$$\frac{190 - 90}{8} = 12.5 \text{ milliamperes.}$$

34. It will be observed that the above calculations have been made with the aid of the assumed static $I_a - V_g$ characteristic. The dynamic characteristic must not be used because the anode circuit load is not the same for all the component variations of effective anode circuit voltage. If, as is the case in all practical circuits, the anode load is an oscillatory circuit tuned to the frequency $\frac{\omega_r}{2\pi}$, its dynamic resistance at any radio frequency $\frac{\omega}{2\pi}$ near but not

necessarily equal to $\frac{\omega_r}{2\pi}$ will be $R_d = \frac{\omega^2 L^2}{R}$ ohms. At radio frequencies considerably larger

than this, e.g. of the order of the second harmonic of $\frac{\omega_r}{2\pi}$, the anode circuit impedance will be negligible owing to the low reactance of the aerial capacitance at these frequencies, while at audio-frequencies the anode circuit simply offers a few ohms resistance. The amplitude of the carrier and side band variations of anode current will therefore be less than that calculated

above in the ratio $\frac{r'_a}{r'_a + R_d}$, where r'_a is the average anode A.C. resistance of the valve for the particular grid swing and is analogous to the conversion resistance of a frequency-changing valve (Chapter XI). As these currents act as the supply to a rejector circuit, the circulating current in the latter will be approximately $\frac{\omega L}{R}$ times as great as the supply current. For example, assuming that the anode load R_d is equal to the average A.C. resistance r'_a of the valve, the carrier-frequency component in our example will be not 70 but 35 milliamperes, and if the magnification $\frac{\omega L}{R}$, of the anode circuit is 50, the aerial current will be 50×35 milliamperes or 1.75 amperes, in the absence of modulation.

“Linear” grid modulation

35. The method of operation just described is of very low efficiency and is rarely if ever used in radio transmitters although it has a useful sphere of employment in radio-frequency transmission on conductive circuits. It has already been shown that if the mean grid potential of an amplifier is maintained at a highly negative value, higher efficiency is possible, and such an amplifier may be subjected to grid bias modulation, comparatively high efficiency being obtainable with little amplitude distortion. It is again convenient to study this mode of operation with respect to a master-oscillator-controlled transmitter, e.g. fig. 20, and to neglect the effects of curvature of the characteristic, the operating conditions being shown in fig. 23. Here the point A is situated at the theoretical “cut-off” voltage and it is assumed to be permissible to allow a small grid current to flow, the maximum positive grid potential being represented

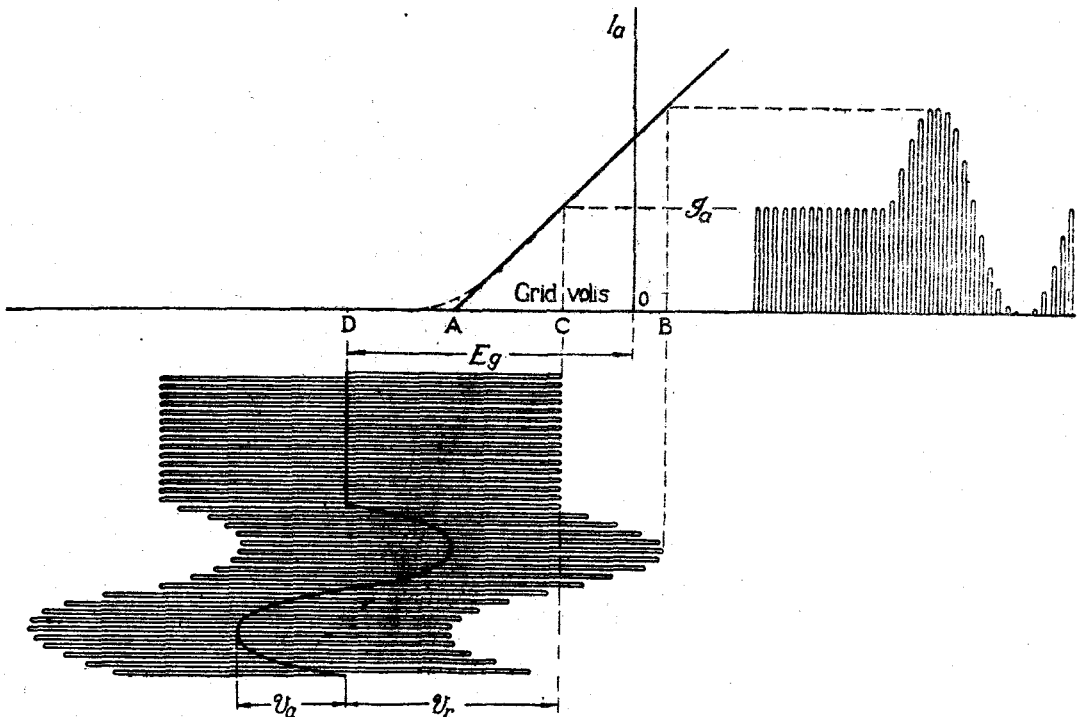


FIG. 23, CHAP. XII.—Linear grid bias modulation.

by the point B while the point C is midway between A and B. The grid bias is sufficiently negative to bring the mean grid potential to the point D where $AD = AC$, and the maximum permissible grid swing is then $2BD$. The amplitude of the radio-frequency input voltage, which is derived from the oscillatory circuit of the oscillator valve, is equal to CD , and the maximum audio-frequency amplitude, derived from the secondary of the microphone transformer, is equal to AD . During periods of peak modulation, the input voltage variations and corresponding changes of anode current are as shown. The audio-frequency envelope of the anode current is a faithful replica of the audio-frequency voltage; the radio-frequency variation of anode current is not sinusoidal, but contains a whole series of radio-frequency components which are not necessarily in harmonic relationship. Owing to the presence of the oscillatory anode circuit, however, only those components which are near the resonant frequency produce appreciable

CHAPTER XII.—PARA. 36

aerial currents. The amplitude of the radio-frequency variation of i_a is $\mathcal{I}_a (1 + \sin \omega_a t)$ and the total variation will be of the form

$$\mathcal{I}_a (1 + \sin \omega_a t) \{ A_1 \sin (\omega_1 t + \varphi_1) + A_2 \sin (\omega_2 t + \varphi_2) + A_3 \sin (\omega_3 t + \varphi_3) \dots \}$$

where no restrictions need be placed on the value of $A_1, A_2, \omega_1, \omega_2$, etc. The important point to observe is that all the frequencies $\frac{\omega_1}{2\pi}, \frac{\omega_2}{2\pi}$, etc., are modulated to a depth of unity, because $(1 + \sin \omega_a t)$ is a common multiplier of all these terms. The first of these, $A_1 \sin (\omega_1 t + \varphi_1)$ is nearly equal to $\frac{1}{2} \sin \omega_r t$, so that the principal effect of the variation of grid voltage is to cause an anode current variation

$$i_m = \frac{r'_a}{r'_a + R_d} \frac{\mathcal{I}_a}{2} (1 + \sin \omega_a t) \sin \omega_r t,$$

the factor $\frac{r'_a}{r'_a + R_d}$ being introduced to allow for the presence of the dynamic resistance in the anode circuit. Thus i_m is modulated to a depth of unity, at the audio-frequency $\frac{\omega_a}{2\pi}$, and the oscillatory current in the aerial circuit will be modulated in like manner.

36. The effect of a slight curvature at the foot of the characteristic will be to introduce some slight degree of amplitude distortion and, for an input voltage ratio $\frac{\mathcal{V}_r}{\mathcal{V}_a} = 2$, to reduce the depth of modulation slightly below unity. An increase in $\mathcal{V}_r$, so that this ratio slightly exceeds 2, may restore the depth of modulation, but may also increase the amount of amplitude distortion. It is important to observe that this method depends for its effectiveness upon the maintenance

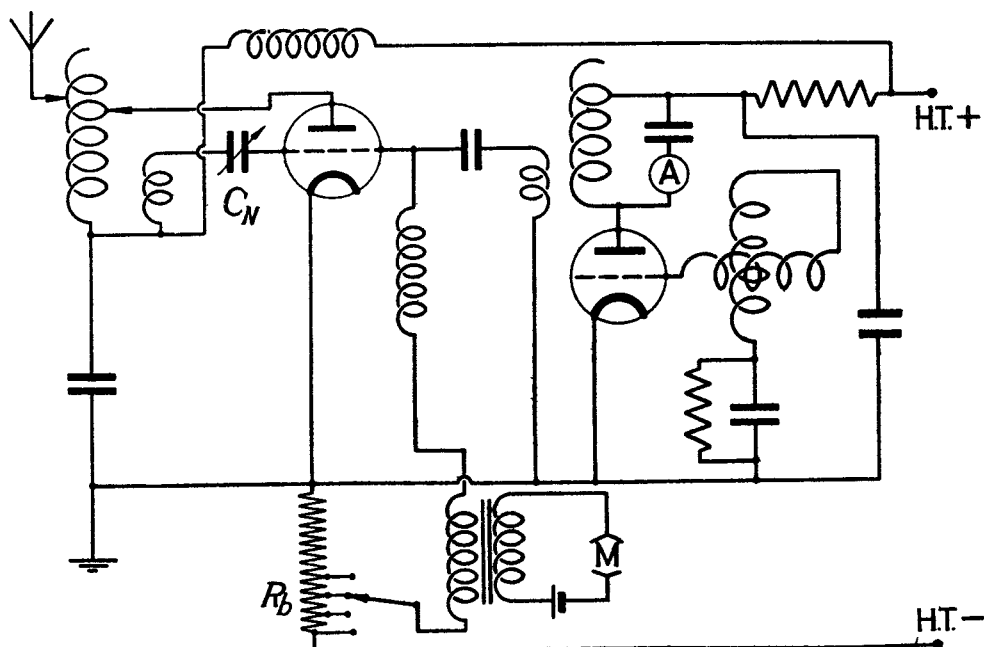
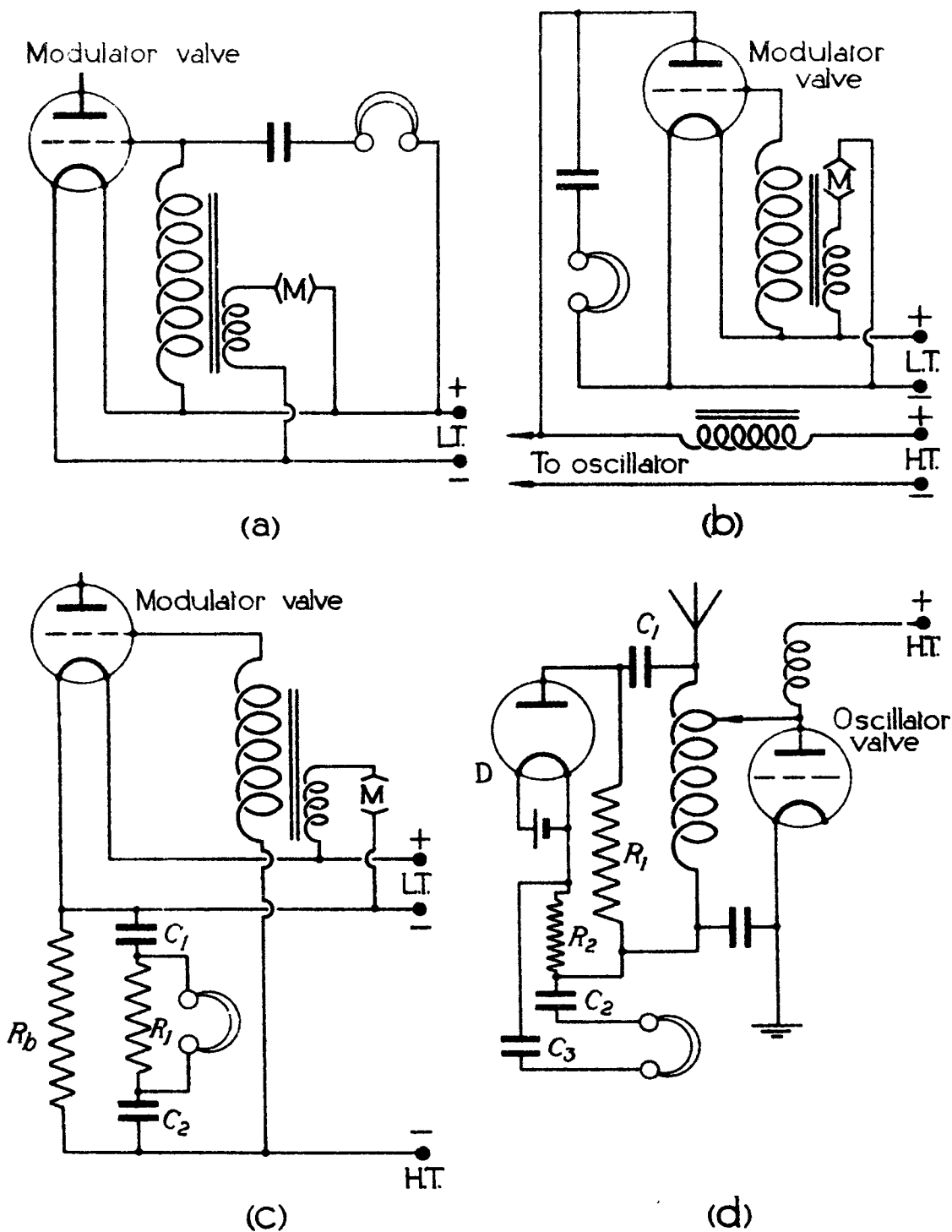


FIG. 24, CHAP. XII.—R/T transmitter with automatic amplifier bias.



SIDE TONE DEVICES

of the mean grid bias at a constant value, irrespective of the instantaneous input grid swing and output power. This requirement forbids the use of a grid condenser and lead for this purpose, for with this device the grid bias varies with the oscillatory output.

37. (i) The circuit diagram of a typical R/T transmitter in which modulation is achieved by the above method is given in fig. 24, in which the neutralizing arrangements for the power amplifying stage are also included. The circuit also shows the method of obtaining automatic grid bias. The resistance R_b is fitted in the H.T. negative conductor and is common to the anode circuits of both valves. A constant P.D. E_g is developed between the ends of this resistance; if I_{pa} and I_o are the mean anode currents of the power amplifier and oscillator valves, respectively, $E_g = R_b (I_{pa} + I_o)$. The direction of current is such that the end of the resistance remote from the filament is at negative potential with respect to the latter, and a number of tappings are provided so that the mean bias may be varied within certain limits. The radio-frequency choke in the grid circuit of the power amplifier maintains the input impedance of this valve at a high value; in its absence the input impedance would be low owing to the comparatively large self-capacitance of the microphone transformer.

(ii) In considering the action of the grid modulator it was assumed that the amplifier valve may be allowed to pass a small grid current. If this is permitted, the efficiency of the amplifier stage will be higher, but the frequency stability of the master-oscillator will be impaired, because the passage of grid current implies a variation in input impedance during each audio-frequency cycle and a corresponding change in the effective resistance of the tuned circuit of the master-oscillator. The bias is therefore generally adjusted to such a value that the grid of the power amplifier is never allowed to reach a positive potential with respect to the filament, and the efficiency under these conditions is only about 40 per cent.

Side tone

38. (i) In order that the operator may have some indication of the correct performance of the transmitter, arrangements are usually made to reproduce his speech in the telephone receivers. This reproduction is called side tone, and its utility depends largely upon the manner by which it is achieved. Thus, if the telephones are connected to the secondary winding of the microphone transformer as in fig. 25a (a very small series condenser being inserted to reduce the secondary load and to limit the telephone current) the side tone serves to indicate first that the microphone and transformer are working properly, and second that the speech level is correct. The latter point is of great importance in transmission from aircraft. This method, however, gives no indication as to the performance of the modulator valve.

(ii) In transmitters using choke control modulation, therefore, it is preferable to connect the telephone receiver and series condenser in parallel with the speech choke; this arrangement serves as a complete check upon the audio-frequency operation of the transmitter. In practice the combination of telephone receiver and series condenser is usually connected directly between the anode and filament of the modulator valve (fig. 25b).

(iii) In transmitters employing automatic grid bias, the arrangement shown in fig. 25c, may be adopted. Here R_b is the resistance which provides the automatic grid bias and corresponds with R_b in fig. 24. The telephones are in effect connected in parallel with the resistance R_b . The condensers C_1 , C_2 are fitted to confine the steady component of anode current to the bias resistance, and by choice of suitable values, in conjunction with the series resistance R_1 , the side tone is maintained at the desired level. The actual values of R_1 , C_1 , C_2 depend upon the power of the transmitter.

(iv) The scheme outlined in fig. 25d serves as a check upon all circuits in the transmitter, for side tone is only produced when the aerial is actually radiating a modulated wave. Here D is a diode rectifier, C_1 a high-voltage condenser of small capacitance, and R_1 a suitable resistance. A small radio-frequency current flows in the path C_1 , R_1 which is parallel with the aerial, and the oscillatory voltage across R_1 causes a rectified current to flow in the resistance R_2 , with which the telephones are effectively in parallel. The condensers C_2 , C_3 serve to isolate the

CHAPTER XII.—PARAS. 39-41

telephones from the aerial circuit. The insulation of the condenser C_1 is of great importance, for this component must withstand the sum of the full oscillatory voltage across the aerial coil and the steady H.T. voltage of the supply.

Frequency distortion

39. In all R/T transmitters, irrespective of the method of modulation, since the side-band frequencies differ slightly from the resonant frequency of the tuned circuits associated with the modulator and amplifier stages, some frequency distortion is bound to occur. The amount of this distortion depends upon the relative magnification of the tuned circuits at the side-band and carrier frequencies; if any circuit has a high magnification the carrier frequency will receive greater amplification than the side-bands and the depth of modulation will be reduced. The problem of avoiding frequency distortion is more difficult on low and medium than on high carrier frequencies, and in any event the side-bands corresponding to high audio-frequencies are attenuated to a greater extent than those corresponding to low audio-frequencies. The obvious remedy is to use tuned circuits having only a low magnification so that the resonant curves are sensibly flat over the frequency band which it is desired to radiate. It is also highly desirable that each resonant curve shall be symmetrical on either side of the resonant frequency so that corresponding side-bands, say 5 kc/s above and below the carrier shall be of equal amplitude.

Effect of unsymmetrical resonance curve

40. (i) Problems involving the relative magnitude and phase of side-bands and carrier are conveniently visualized by means of vector diagrams. The principle will first be exhibited as applied to an oscillatory current of carrier amplitude $\mathcal{J}_c$ and frequency f_c modulated sinusoidally to a depth of unity. Let $\mathcal{J}_H$ be the amplitude of a side-frequency $f_c + f_a$ and $\mathcal{J}_L$ be the amplitude of the corresponding side-frequency $f_c - f_a$. Then the carrier may be represented by a vector rotating in the positive direction f_c times per second, the higher side-frequency by a vector rotating $f_c + f_a$ times per second and the lower side-frequency by a vector rotating at $f_c - f_a$ times per second. With reference to the carrier vector $\mathcal{J}_c$, therefore, $\mathcal{J}_H$ may be considered to rotate at a frequency $+f_a$ and $\mathcal{J}_L$ at the frequency $-f_a$. The negative sign may now be considered to signify contrariety of direction and if the carrier is considered to be stationary, $\mathcal{J}_H$ rotates at the frequency f_a in the counter-clockwise (positive) direction, while $\mathcal{J}_L$ rotates at the frequency f_a in the clockwise (negative) direction. The amplitude of the vector sum of the two side-frequencies may be denoted by $\mathcal{J}_s$ and the amplitude of the modulated oscillation at various intervals during a single audio-frequency cycle may then be derived as shown in fig. 26b. It is seen that if the side-bands are of equal amplitude and in phase opposition as is normally the case, $\mathcal{J}_s$ is always either in phase or in anti-phase with $\mathcal{J}_c$ and the amplitude of the resultant modulated oscillation at any instant during the audio-frequency cycle is obtained by algebraic addition of $\mathcal{J}_c$ and $\mathcal{J}_s$.

(ii) Now let us consider the effect of unequal amplification of upper and lower side-bands. Suppose that $\mathcal{J}_H$ is not equal to $\mathcal{J}_L$, although they are still in phase opposition. The conditions are now as illustrated in fig. 26c where it is seen that the vector sum of $\mathcal{J}_H$ and $\mathcal{J}_L$ varies in phase with respect to the carrier $\mathcal{J}_c$, the resultant, $\mathcal{J}$, being the vector sum of $\mathcal{J}_s$ and $\mathcal{J}_c$ and not the algebraic sum. The vector $\mathcal{J}_s$ sometimes leads and sometimes lags upon $\mathcal{J}_c$, so that with respect to the carrier frequency, the resultant oscillation undergoes a cyclical change of phase, and is modulated both in phase and amplitude.

Intermodulation by non-linear microphone

41. When a single sinusoidal sound wave represented by $p = P_o \sin \omega t$ impinges upon the diaphragm of a carbon microphone, it was shown in paragraph 10 that the equivalent voltage generated is

$$v_m = -I_m R_m \left\{ \frac{P_o}{P} \sin \omega t - \left(\frac{P_o}{P} \right)^2 \sin^2 \omega t \dots \dots \dots \right\}$$

It is of interest to examine the effect produced when two or more sinusoidal sound waves are

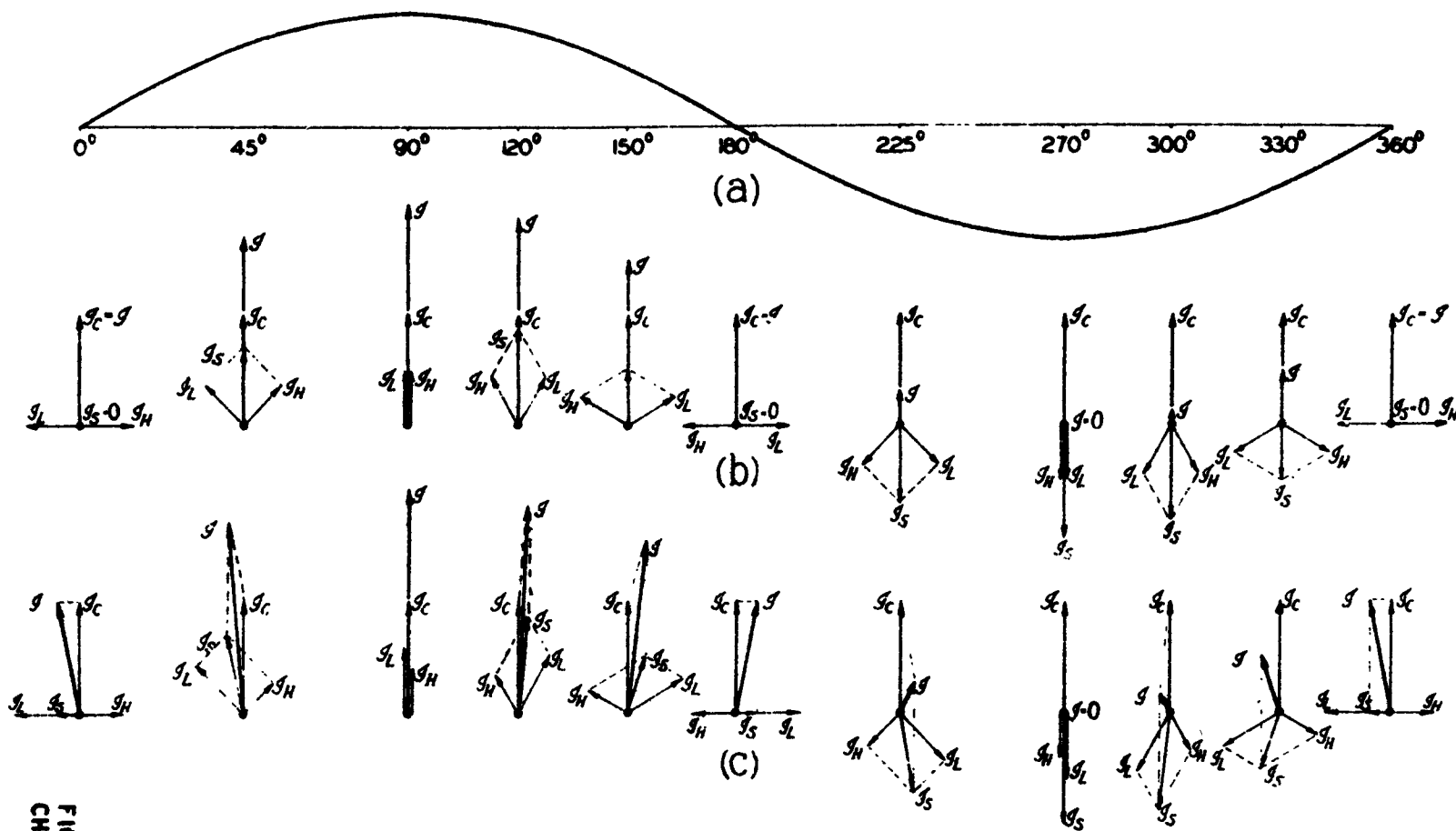


FIG. 26
CHAP. XII

VECTORS REPRESENTING CARRIER & SIDE-BANDS

applied simultaneously. Let two sinusoidal sound waves, of amplitudes P_1 and P_2 and frequencies $\frac{\alpha}{2\pi}$, $\frac{\beta}{2\pi}$ be simultaneously applied, then the total sound pressure on the diaphragm will be $P_1 \sin \alpha t + P_2 \sin \beta t$. The voltage generated will therefore be proportional to

$$\frac{P_1 \sin \alpha t + P_2 \sin \beta t}{P} = \frac{(P_1 \sin \alpha t + P_2 \sin \beta t)^2}{P^2}$$

or

$$\begin{aligned} & \frac{P_1}{P} \sin \alpha t + \frac{P_2}{P} \sin \beta t - \left(\frac{P_1}{P}\right) \sin^2 \alpha t - \\ & \frac{2P_1 P_2}{P^2} \sin \alpha t \sin \beta t + \left(\frac{P_2}{P}\right)^2 \sin^2 \beta t \end{aligned}$$

The terms containing $\sin^2 \alpha t$ and $\sin^2 \beta t$ represent second harmonic distortion as already shown. It will be observed however that the voltage now contains a modulation product $\frac{2P_1 P_2}{P^2} \sin \alpha t \sin \beta t$.

$$= \frac{P_1 P_2}{P^2} \cos (\alpha - \beta) t - \cos (\alpha + \beta) t.$$

The result is in fact that owing to the non-linear response of the microphone the two sound waves modulate each other. When they are merely two speech waves of different frequencies and of small amplitudes so that $\frac{P_1 P_2}{P^2}$ is of the same order as $\left(\frac{P_1}{P}\right)^2$ and $\left(\frac{P_2}{P}\right)^2$ the result of this intermodulation is no more serious than the introduction of second harmonic distortion, but if one of them, say P_2 , represents a very loud interfering noise such as that caused by the airscrew in an aeroplane, it may be many times greater than the average pressure due to speech, and the sum and difference terms will give rise to considerable loss of intelligibility. The actual effect is, of course, much more complex than when only two sinusoidal components are considered, but in general the result at the receiver is to impart to the speaker's voice a peculiar tremolo character which is unmistakable. It must be noted also that this effect is additional to the interference caused by the transmission of the frequency $\frac{\beta}{2\pi}$ in the ordinary manner. Intermodulation can only be completely overcome by the use of a microphone having a perfectly linear characteristic. Either moving coil or condenser microphones would probably effect a considerable improvement, but only at the cost of a considerable increase in the weight, volume and complexity of the installation. The latter considerations dictate the use of carbon microphones in most aircraft installations, but intermodulation by engine noises is reduced to a minimum by careful design of the mask microphone.

Suppressed carrier and single side-band telephony

42. It has been shown that in the ordinary amplitude-modulated R/T transmitter, the power carried by the carrier frequency is at least equal to and may be very much greater than the power carried by the side-band frequencies, although only the latter convey the desired intelligence. The carrier frequency is required in the first instance to produce the side-bands by interaction with the audio-frequency input, and it is necessary at the receiver in order to avoid distortion. Once the desired side-bands have been obtained there is however no reason why the carrier frequency should not be eliminated in the amplifier stages of the transmitter, and reintroduced at the receiver. The transmitter is then required to handle considerably less power and may also operate at a higher efficiency. One method of producing the desired side-bands and eliminating the carrier is by means of the balanced modulator shown in fig. 27 which is based upon the circuit used for grid modulation. The two triodes T_1 , T_2 are of similar (preferably identical) characteristics, and the input and output circuits are arranged symmetrically with respect to them. In the absence of an audio-frequency input the grid-filament potential is varied in accordance with the radio-frequency input derived from the master-oscillator. To

CHAPTER XII.—PARA. 43

this voltage the grid-filament paths of the valves are in parallel and the output circuit is in push-pull, so that the radio-frequency voltages developed in the two halves, L_1 , L_2 , of the inductance in the output circuit are in opposition. No power is developed in the oscillatory circuit $L_a C_a$ under these conditions. To the audio-frequency voltage supplied by the microphone transformer, both input and output circuits are in push-pull, and the grid-filament voltages of the two valves are in antiphase. The result of the simultaneous application of audio-frequency and radio-frequency input voltages is to develop, in the output circuit, an oscillatory E.M.F. corresponding to the beats between side-bands. Thus, if the audio-frequency input voltage is a sinusoidal wave-form, the output voltage is as shown in fig. 6c, taking the form of heterodyne beats, and it must be remembered that two beats are formed for each audio-frequency cycle. This voltage may now be amplified as found desirable, the final output circuit being the transmitting aerial.

43. Neglecting any distortion due to the medium of propagation, the aerial current at the receiver will have precisely the same wave form, and on rectification this would give an audio-frequency output of double the original modulating frequency; if this disadvantage is to be

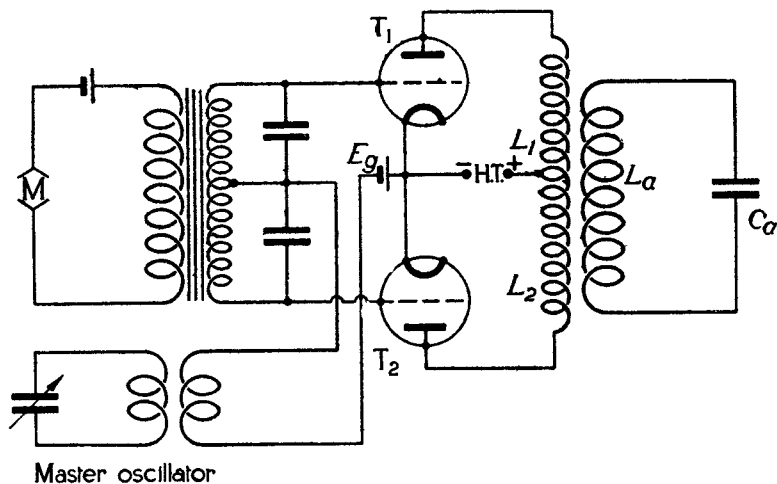


FIG. 27, CHAP. XII.—Balanced modulator.

avoided, the carrier must be reintroduced before the wave is rectified. The mere replacement of the carrier, however, is not in itself sufficient to ensure that the detector output will possess a wave-form corresponding with the original. Referring again to fig. 6, it is seen that the sinusoidal envelope of the amplitude modulated wave is entirely dependent upon the change of phase between adjacent beats. If the carrier is not reintroduced in its correct phase, the rectified output of the detector will be distorted. Fig. 28a has been developed to show what happens in an extreme case, when the reintroduced carrier is 90° out of phase with respect to the original carrier. A single "beat" is drawn in light line and the replaced carrier in heavy line. The respective amplitudes are such that if the relative phase were correct the resultant would be modulated to a depth of unity. Actual point-to-point addition of the two waves, as in fig. 28b, shows however that the resultant of the two waves is modulated to a depth of only about 15 per cent. and the shape of the envelope indicates that rectification would still give an output at double the original frequency. Thus the reintroduction of the carrier with a phase displacement of this magnitude would reduce the strength of the rectified signal without reducing the distortion. From the practical point of view, this rules out the possibility of replacing the carrier merely by using, as part of the receiving equipment, a separate heterodyne tuned to the carrier frequency, although with the introduction of automatic tuning control there is some hope of a solution of the problem. At the present time, however, suppressed carrier working is rarely if ever adopted

44. Since each of the two side-bands contains the whole of the intelligence to be transmitted it is possible to convey the desired signal by means of only one side-band, eliminating the carrier by means of the balanced modulator and the unwanted side-band by means of a band-pass filter of the Campbell-Zobel type. The carrier is then reintroduced at the receiver by means of a separate heterodyne, which beats with the frequencies embraced by the single side-band. After rectification the resulting difference frequencies reproduce the original audio frequencies, and although some distortion is introduced, the effect of the relative phase shift of replaced carrier and single side-band is not so serious as when the carrier only is suppressed, provided that the amplitude of the heterodyne oscillation is much larger than that of the received signal. Single side-band working is not more economical than suppressed carrier working, but it reduces the frequency band occupied by a given transmitter by one-half. The additional complexity

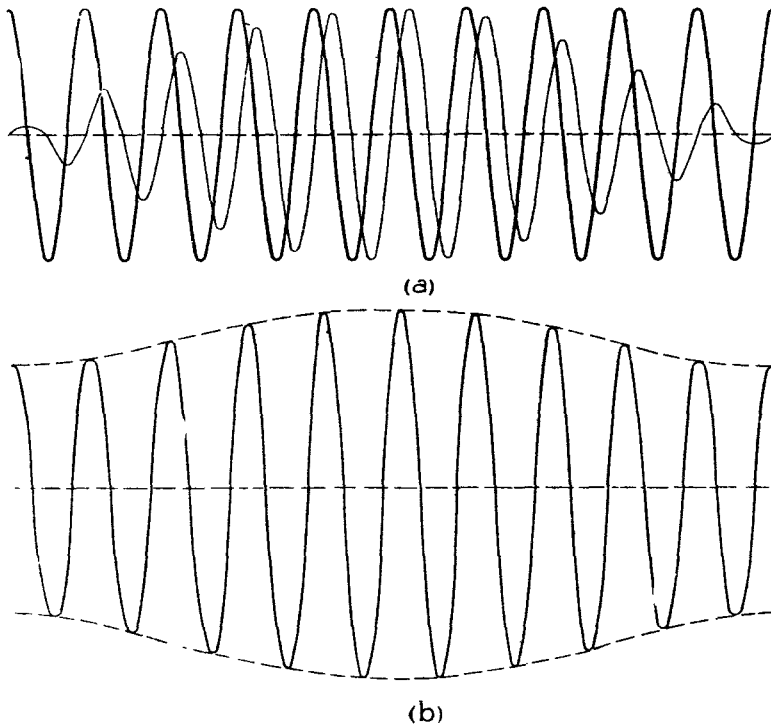


FIG. 28, CHAP. XII.—Effect of phase displacement of carrier.

renders it suitable only for ground-to-ground communication between high-power stations operating on a single fixed-frequency channel. The stringent operating requirements are more easily met on low and medium than on high frequencies, although progress is being made in its application to the latter.

R/T RECEPTION

45. In principle, the circuits used for reception of amplitude-modulated radio-telephonic signals are very similar to those used for telegraphic reception. The arrival of modulated electromagnetic waves at the receiving aerial sets up an induced E.M.F. of identical wave form, a corresponding radio-frequency current being established in the aerial circuit. The process of detection in this case is the reverse of the modulation process at the transmitter, for the detector

CHAPTER XII.—PARA. 46

is required to give an audio-frequency output having a wave-form identical with that of the envelope of the modulated signal. Radio-frequency amplification may be and in practice generally is employed in order to achieve the highest degree of selectivity possible (space, weight and cost of the receiver being taken into account), and also in order that the detection process shall be carried out efficiently. After this, audio-frequency amplification is employed to bring the final reproduction to a level sufficiently high to be audible over the local noise, either head telephones or loud speaker being employed to convert the electrical power output of the final amplifier stage into sound waves. Whereas for C.W. reception the sensitivity of the receiver is of chief importance, neither amplitude, phase or frequency distortion being detrimental to efficient reception, in an R/T receiver fidelity is an important consideration, i.e. the audible output of the receiver must be as faithful a reproduction as possible of the original sound vibrations.

Selectivity

46. (i) To a first approximation, distortion of the radio-frequency wave-form is of no importance provided the original shape of the modulation envelope is preserved; to achieve this the radio-frequency circuits preceeding the detector stage must give equal magnification of the

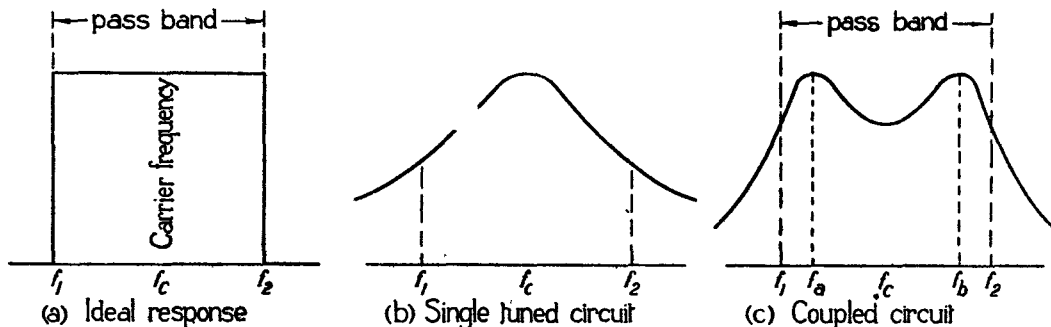


FIG. 29, CHAP. XII.—Ideal and practicable response characteristics.

whole range of frequency embraced by the carrier and side-bands, and an ideal response characteristic for this portion of the receiver would have the form given in fig. 29a, in which all frequencies between f_1 and f_2 receive equal amplification while frequencies outside this range are attenuated to zero. The frequency range between f_1 and f_2 is called the "pass-band"

(ii) If the receiver contains only a single tuned circuit, its frequency response will be governed by the resonance curve of the tuned circuit and will resemble fig. 29b. Since all frequencies above and below the carrier frequency f_c are attenuated to some extent, the receiver cannot strictly be said to possess any pass-band whatever. An approach to the ideal may be obtained by the use of coupled circuits of the types described in Chapter VI, which have a frequency response similar to that shown in fig. 29c. The pass-band of such a receiver is usually taken as the region within which the height of the curve is equal to or greater than its height at the carrier frequency f_c . This is 1.414 times the frequency separation of the "peaks" of the resonance curve. For telephony of commercial quality a pass band of some 5 kc/s is sufficient, but for faithful reproduction of music and particularly of noises such as the rattling of keys, all frequencies up to about 15 kc/s are required and the pass-band of the receiver should be some 30 kc/s wide. The shape of the resonance curve of two coupled circuits depends upon the coupling factor, k , and upon the circuit magnification of each member. If the magnification of the primary circuit is x_p , that of the secondary circuit x_s , and both are tuned to the frequency f_c , it is shown in Chapter VI that provided $k^2 > \frac{1}{x_p x_s}$ the resonance curve possesses two peaks,

the frequencies at which they occur being

$$f_a = \frac{f_c}{\sqrt{1+k}}$$

$$f_b = \frac{f_c}{\sqrt{1-k}}$$

while a trough occurs at the frequency f_c . It must be observed that f_a does not correspond with f_1 , nor f_b with f_2 , in fig. 29c. The peak separation is $f_b - f_a$ and the pass-band is $f_2 - f_1 = \sqrt{2}(f_b - f_a)$.

47. The relative magnification at the frequencies f_a and f_c (or f_b and f_c) depends upon the circuit magnification. In order to obtain an approach to the ideal response characteristic, fig. 29a, it is obvious that the peaks of the resonance curve must be nearly but not quite suppressed. The width of the pass-band depends chiefly on the coupling factor k , while the uniformity of response within the band depends upon k , χ_p , and χ_s . The degree of discrimination against frequencies just outside the pass-band also depends upon k , χ_p , and χ_s . It is not under

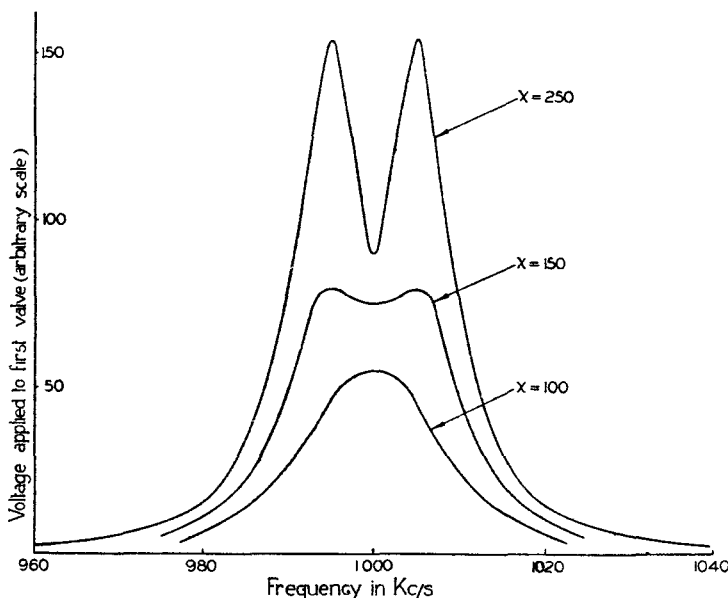


FIG. 30, CHAP. XII.—Effect of varying resistance of coupled circuit.

separate control. In showing the effect of varying the magnification and coupling factor it is convenient to assume that $\chi_p = \chi_s = \chi$. The effect of varying k , while keeping χ constant, has been shown in Chapter VI. In fig. 30 the coupling factor is maintained at a constant value ($k = .01$) and the circuit magnification is varied by adjustment of the resistance. It is seen that if a high value of χ is adopted, the response characteristic possesses pronounced double peaks, so that after rectification the higher audio-frequencies will be exaggerated and the reproduction will appear to be highly pitched. On the other hand, very low values of χ cause the lower audio-frequencies to receive greater magnification than the higher, and the reproduction will appear "woolly". The curve marked $\chi = 150$ represents a very close approach to the ideal, giving a very even response over a band width of about 13 kc/s and a high degree of discrimination against the frequencies outside the pass-band.

CHAPTER XII.—PARA. 48

48. The effect of simultaneously increasing k , in order to obtain a wider pass-band, while at the same time reducing χ (by an increase of resistance) in order to suppress the double peaks, may be seen from fig. 31. A wide band having an even response can only be achieved at the expense of a reduction in output voltage. The effective magnification of the whole arrangement is $\frac{1}{2} \sqrt{\chi_p \chi_s}$, e.g. if $\chi_p = \chi_s$, the voltage gain is only one-half that which would be given by a single circuit of the same magnification. This is the price which must be paid in return for the "band-pass" effect. The curves of figs. 30 and 31 were calculated on the assumption that the circuits are coupled by mutual inductance, but the above deductions are applicable to all forms of reactive coupling. Resistance coupling is seldom if ever deliberately adopted in receiving circuits. The term "band-pass filter" is used in broadcast receiver practice to describe an arrangement of two coupled circuits in which both members are simultaneously adjusted to the same frequency, e.g. by using equal inductances in each member and inter-connecting the

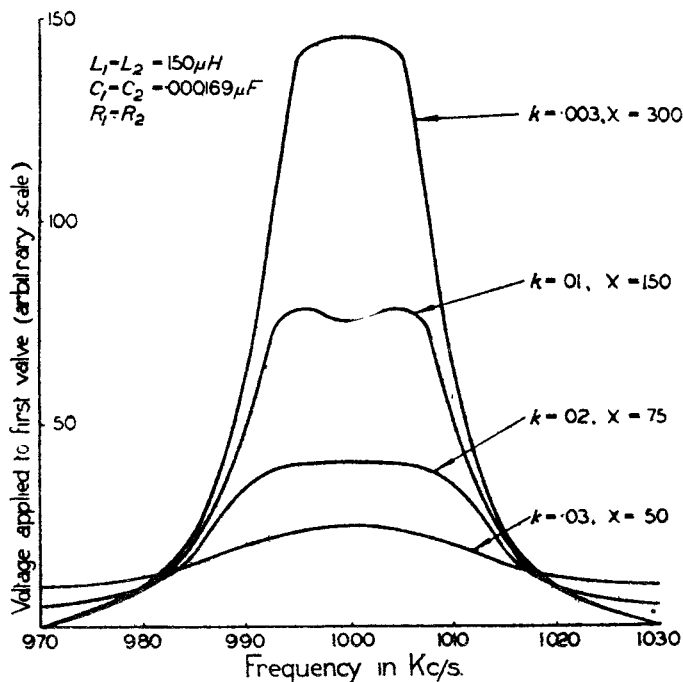


FIG 31, CHAP. XII.—Effect of simultaneous variation of magnification and coupling factor.

tuning condensers in such a manner that both are adjusted to the desired capacitance by a single tuning control. In choosing the type of coupling reactance an important consideration is the variation in width of pass-band as the tuning is varied. Thus if mutual inductive coupling is employed, since the width of the pass-band is $f_2 - f_1 = \sqrt{2} (f_0 - f_s) = \sqrt{2} f_c k$, the pass-band increases with the frequency to which the individual circuits are tuned. Assuming that the maximum capacitance of the condenser is ten times the minimum (including of course the distributed capacitance of the coil and winding) it is seen that if the band width is adjusted to 10 kc/s at the lowest frequency of operation, for which the tuning condenser is adjusted to its maximum value, the band width at the highest frequency will be $\sqrt{10} \times 10$ kc/s or 31.6 kc/s.

This follows from the relation $f_c = \frac{1}{2\pi \sqrt{LC}}$, because if L is constant $f_c \propto \frac{1}{\sqrt{C}}$. On the other

hand, if auto-capacitive coupling is employed, $k = \frac{C}{C + C_m} \doteq \frac{C}{C_m}$, because C_m is always very much larger than C . The band width is therefore directly proportional to $\sqrt{C}$ and increases as the operating frequency is decreased. In practical circuits of this type an attempt is frequently made to overcome this disadvantage by a combination of different forms of coupling.

Amplitude distortion

49. (i) While the radio-frequency circuits may cause some degree of frequency distortion, the first source of appreciable amplitude distortion is the rectifying valve, by which the modulated wave is resolved into its radio-frequency and audio-frequency components and the former discarded. It has already been shown that for small amplitudes of input voltage all practical rectifiers give an output which is proportional to the square of the input (except in the case of heterodyne reception). The "square law" rectification of a completely modulated wave by means of a diode will result in an anode current variation somewhat as depicted in fig. 32, in

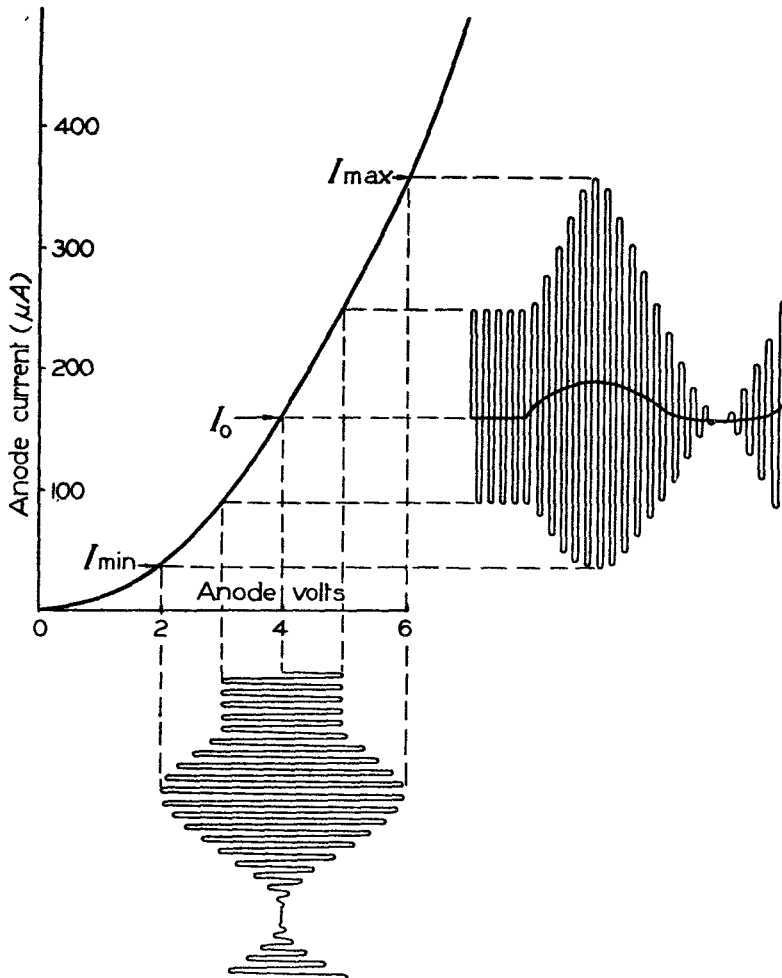


FIG. 32, CHAP. XII.—Square law rectification of modulated wave.

CHAPTER XII.—PARA. 49

which the wave form of the audio-frequency component is shown by a heavy line ; its shape obviously differs from that of the modulation envelope of the applied voltage, i.e. amplitude distortion has taken place.

(ii) The degree to which distortion is introduced is easily calculated from the known equation of the diode characteristic.

$$\begin{aligned} \text{Let} \quad i_a &= g v_a^2 \\ \text{and} \quad v_a &= E_a + \mathcal{V}_r (1 + \sin \omega_a t) \sin \omega_r t \end{aligned}$$

Here E_a is a steady anode voltage and $\mathcal{V}_r (1 + \sin \omega_a t) \sin \omega_r t$ a radio-frequency voltage modulated to a depth of 100 per cent. Then

$$\begin{aligned} i_a &= g \{ E_a + \mathcal{V}_r \sin \omega_r t (1 + \sin \omega_a t) \}^2 \\ &= g \{ E_a^2 + 2 E_a \mathcal{V}_r \sin \omega_r t + \mathcal{V}_r^2 \sin^2 \omega_r t \\ &\quad + [2 E_a \mathcal{V}_r \sin \omega_r t \sin \omega_a t] \\ &\quad + [2 \mathcal{V}_r^2 \sin^2 \omega_r t \sin \omega_a t] \\ &\quad + [\mathcal{V}_r^2 \sin^2 \omega_r t \sin^2 \omega_a t] \} \end{aligned}$$

Only the terms containing $\omega_a t$ are able to contribute to the audio-frequency output and these are enclosed in square brackets in the final equation. They are

$$(a) 2E_a \mathcal{V}_r \sin \omega_r t \sin \omega_a t.$$

This is equivalent to

$$E_a \mathcal{V}_r \{ \cos (\omega_r - \omega_a) t - \cos (\omega_r + \omega_a) t \}$$

and represents anode current components corresponding to the side-band frequencies.

$$(b) 2 \mathcal{V}_r^2 \sin^2 \omega_r t \sin \omega_a t.$$

This can be written

$$\mathcal{V}_r^2 (1 - \cos 2 \omega_r t) \sin \omega_a t$$

and yields a component of frequency $\frac{\omega_a}{2\pi}$ and amplitude $g \mathcal{V}_r^2$ together with other components of frequency $\frac{2\omega_r \pm \omega_a}{2\pi}$, which are of no immediate interest.

$$(c) \mathcal{V}_r^2 \sin^2 \omega_r t \sin^2 \omega_a t.$$

This is equivalent to

$$\begin{aligned} &\frac{\mathcal{V}_r^2}{4} (1 - \cos 2\omega_r t) (1 - \cos 2\omega_a t) \\ &= \frac{\mathcal{V}_r^2}{4} \left\{ 1 - \cos 2\omega_r t - \cos 2\omega_a t + \cos 2\omega_a t \cos 2\omega_r t \right\} \end{aligned}$$

yielding certain radio-frequency components, together with an audio-frequency component of amplitude $g \frac{\mathcal{V}_r^2}{4}$ and frequency $\frac{2\omega_a}{2\pi}$. The latter component, being of twice the frequency of modulation, is referred to as second harmonic distortion.

(iii) The total audio-frequency variation of anode current, to which the telephones will respond, is therefore

$$i_a' = g \left\{ \mathcal{V}_r^2 \sin \omega_a t - \frac{\mathcal{V}_r^2}{4} \cos 2\omega_a t \right\},$$

and the square law rectification of a completely modulated wave is shown to result in the introduction of second harmonic distortion, the amplitude of the harmonic being 25 per cent. of the fundamental component. For any other depth of modulation, K , the percentage of second harmonic distortion is $\frac{K}{4} \times 100$. The mean increase of anode current during rectification is

also equal to $g \frac{K V_g^2}{4}$ so that if the mean increase is found the amplitude of the second harmonic is known. Applying the method given in Chapter IX, therefore, the percentage of second harmonic distortion is found from the formula.

$$\left. \begin{array}{l} \text{Percentage of second} \\ \text{harmonic distortion} \end{array} \right\} = \frac{\frac{1}{2} (I_{\max.} + I_{\min.}) - I_0}{I_{\max.} - I_{\min.}} \times 100$$

the notation used being shown in fig. 32.

This formula may be used to estimate the second harmonic distortion introduced by the audio-frequency stages, even though the valve characteristic is not truly parabolic. It is generally accepted that if the second harmonic distortion is less than 5 per cent. it is hardly perceptible to the ear, that in musical reproduction, up to 10 per cent. may be tolerated, and that for commercial or service telephony up to 15 or 20 per cent. is permissible.

Triode rectification—general

50. For a small input grid swing, say up to about .5 volt, a triode operates approximately as a square law rectifier, no matter whether the curvature of the $I_g - V_g$ or $I_a - V_g$ curve is employed, and the sound output is considerably distorted unless the average depth of modulation is low. When possible, however, the detector of an R/T receiver is operated with a grid swing of at least 3 volts (unmodulated carrier), i.e. an input voltage of about 1 volt R.M.S. Under these conditions, provided that the depth of modulation does not exceed about 80 per cent., the audio-frequency output current is practically proportional to the input voltage and the rectifier is said to be linear.

Anode circuit linear rectification

51. (i) The dynamic characteristic of a typical detector valve, with an anode load of 100,000 ohms, is given in fig. 33. Assuming that provision is made for operation with a carrier amplitude not exceeding 4.5 volts, and 100 per cent. modulation, the H.T. voltage in this particular case has been so chosen that the anode current cut-off occurs at a grid voltage of -9 volts, and the mean bias is adjusted to this value in the usual manner. The input voltage shown has a carrier amplitude of 4.5 volts but is modulated to a depth of only 80 per cent. The variation of anode current consequently occurs over that portion of the characteristic which is practically straight; a rectifier operated under conditions such as these is usually referred to as a linear anode bend detector. The envelope of the anode current variation is to all intents and purposes a copy of the envelope of the input voltage, while the audio-frequency component of anode current, shown in heavy line, closely follows the wave form of the original sound wave.

(ii) Under actual operating conditions, the anode circuit load resistance is shunted by a condenser having a comparatively low reactance at the carrier and side-band frequencies. If this condenser is omitted, the detector valve will impose a heavy damping upon the input circuit owing to the miller effect. On the other hand, if its reactance at the carrier frequency is very small compared with the anode A.C. resistance of the valve, the damping due to the valve will be negligible, but the radio-frequency variations of anode current will correspond with the static and not with the dynamic characteristic. Since the former has a greater curvature than the latter the distortion will be more pronounced. This condenser is, in effect, the reservoir condenser of the rectifier, and the determination of its capacitance C_0 is subject to the somewhat conflicting considerations mentioned above.

CHAPTER XII.—PARA. 52

52. Stated more fully, these considerations are :—

- (i) the reactance, X_o , of the reservoir condenser at the highest audio-frequency must be very much larger than the load resistance, R_o ,
- (ii) its reactance at carrier frequency must be small compared with the load resistance,
- (iii) the time constant, $C_o R_o$ of the anode circuit must be small in order that the output may follow the variation of input voltage as closely as possible. As R_o must be large in order to give the desirable linearity of dynamic characteristic, this requirement calls for a small value of C_o .
- (iv) the input admittance of the valve must be as low as possible.

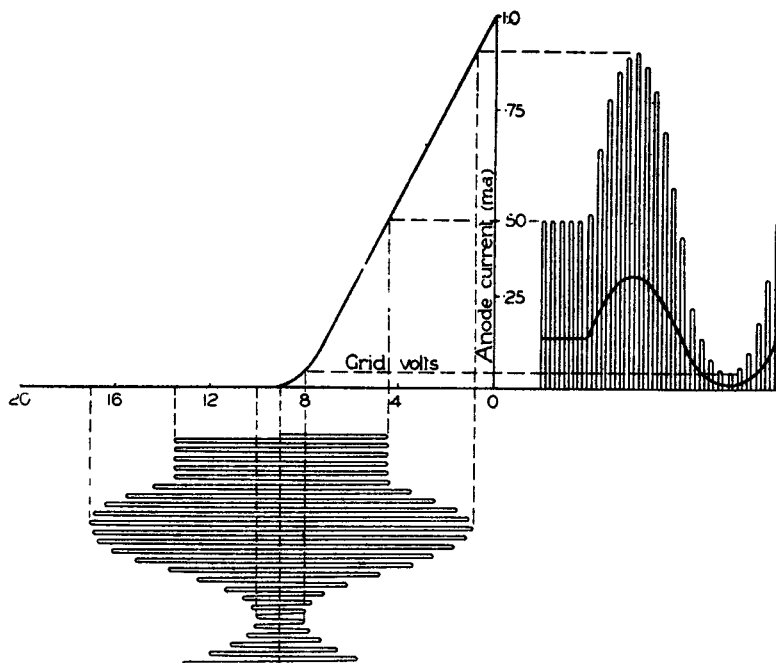


FIG. 33, CHAP. XII.—Anode circuit linear rectification.

Applying these considerations to a practical example, let the highest audio-frequency be 8,000 cycles per second ($\omega = 5 \times 10^4$) and let us assume that to meet requirement (i) a ratio $\frac{X_o}{R_o} = 5$ will be satisfactory. Then if $R_o = 100,000$ ohms, $X_o = 5 \times 10^5$,

$$\begin{aligned} \frac{1}{\omega C_o} &= 5 \times 10^5 \\ C_o &= \frac{1}{5 \times 10^4 \times 5 \times 10^5} \text{ farad.} \\ &= .00004 \mu\text{F.} \end{aligned}$$

To meet requirement (ii) the reactance of the condenser C_o , at the carrier frequency, must be small compared to R_o . For example, if the carrier frequency be 796 kc/s ($\omega = 5 \times 10^6$) then the reactance of a capacitance of $.00004 \mu\text{F}$ is 5,000 ohms which is satisfactory. The time constant will be $C_o R_o = .00004 \times 10^{-6} \times 10^5 = .000004$ second. Since this time is not very

much greater than the duration of one radio-frequency cycle, the envelope of the anode current variation will follow that of the grid voltage very closely and requirement (iii) is satisfied. The effect of the time constant is shown qualitatively in fig. 34. For simplicity, the grid-filament voltage is assumed to vary in amplitude in an abrupt manner (fig. 34a). If the time constant is infinitely small, the anode current will vary in such a manner that the telephone current will reproduce exactly the envelope of the grid filament voltage, taking the form indicated in fig. 34b. With a finite time constant, the anode current will not vary in precisely the same manner as the envelope of the grid-filament voltage. Provided $C_o R_o$ is comparatively small, the anode current will vary somewhat as in fig. 34c, while a further increase in $C_o R_o$ will result in the anode current variation shown in fig. 34d. For a given load resistance R_o , therefore, it is obvious that the value of C_o must be a compromise between the values indicated by requirements (i) and (iv) since the latter indicates the largest practicable value of C_o .

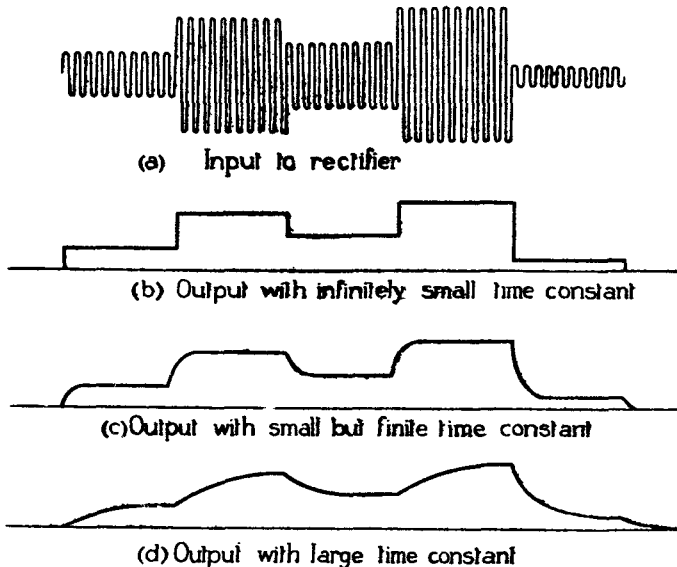


FIG. 34, CHAP. XII.—Effect of detector time constant.

53. The operating potentials of the detector valve are specified on the assumption that a definite grid swing is available. The radio-frequency stages preceding the detector must be designed to give this input voltage at the detector, with a given field strength, and the appropriate aerial. Thus if the amplitude of the grid-filament voltage at the detector is to be 4.5 volts, and the appropriate aerial has an effective height of 2 metres, while the receiver is to give the specified output with a field strength of 45 micro-volts per metre, the total voltage gain of the radio-frequency stages must be $\frac{4.5}{90 \times 10^{-6}} = 5 \times 10^4$. Three stages, each having a gain of about 37, are therefore required, or assuming that a circuit magnification of 20 may be achieved before applying the signal to the first valve, two stages each giving a voltage amplification of 50. This gain can easily be realized at frequencies below about 1,000 kc/s. Signals of considerably lower field strength will also be received, though at reduced strength. If however the field strength is sufficiently great to give a detector input greatly exceeding 4.5 volts (with the valve represented in fig. 33), the detector will be overloaded and will give rise to distortion. The remedy for this is to provide some form of gain control in the radio-frequency amplifier stages and thus limit the detector input to that for which it is designed.

CHAPTER XII.—PARA. 54

Grid circuit (linear) rectification

54. This method of detection is used to a greater extent than anode circuit rectification because, for a given input, it is capable of delivering a somewhat greater output with less distortion. The circuit used is identical with that used for the cumulative grid rectification of C.W. or I.C.W. signals, and is given in Chapter IX. The values of grid condenser and grid leak resistance are, however, primarily chosen with a view to avoiding distortion, rather than for maximum signal strength. The detection of a modulated signal is very similar to that of a C.W. signal with separate heterodyne and may be explained with reference to figs. 35 and 36. The action in the grid circuit is shown in fig. 35; the input signal voltage is assumed to be similar to that of fig. 33, and its original envelope is shown in heavy line. If the grid were connected

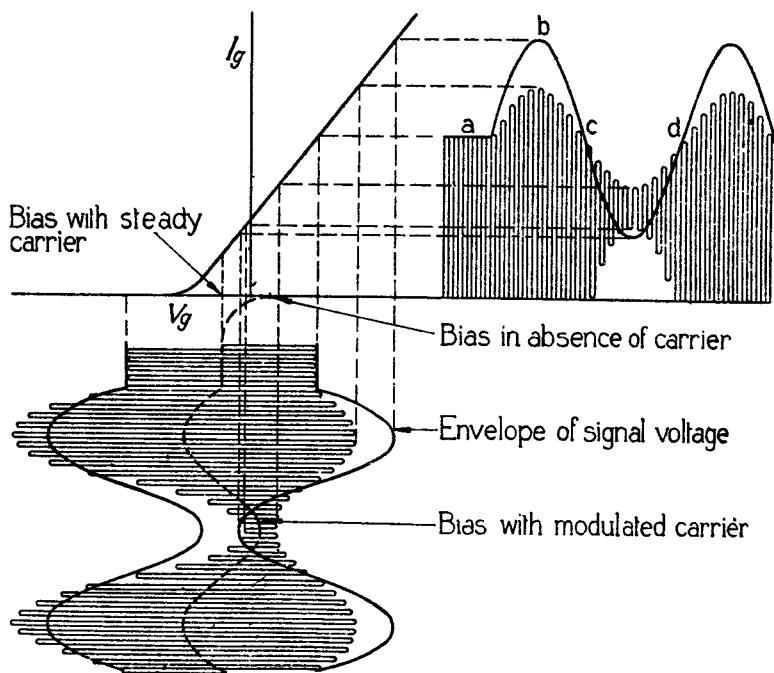


FIG. 35, CHAP. XII.—Grid circuit linear rectification; action at grid.

directly to the filament, the envelope of the resulting variation of grid current would be as indicated by the curve *a, b, c, d*. Since, however, the grid current must return to the filament via the grid leak resistance, the grid voltage varies in a manner corresponding with the changes of anode current. The actual variations of radio-frequency input voltage and grid current are therefore as shown in light line, and the audio-frequency variation in grid bias in dotted line. The latter has a wave-form similar to that of the original modulation envelope. The effect of the total variation of grid voltage upon the anode current is shown in the following diagram (fig. 36). The anode current varies at the radio frequency, and its mean value also varies in accordance with the mean variation of grid bias, i.e. at audio frequency. The radio-frequency variations are by-passed by a suitable condenser, and are not of immediate concern. The audio-frequency variation may be caused to operate the telephone receivers either directly or, as is more usual, after one or more stages of audio-frequency amplification.

Values of grid leak and condenser

55. In a grid circuit rectifier the grid condenser is in effect the reservoir condenser of the rectifier and the actual rectifier load is the grid leak resistance. The voltage across the rectifier condenser can only decrease at the rate allowed by the load resistance, and, as in anode circuit rectification, the condenser voltage can follow the modulation envelope only if the time constant is small. The fluctuation of the modulation envelope is obviously more rapid for high audio frequencies than for low, and for deep than for shallow modulation; for any given audio frequency $\frac{\omega_a}{2\pi}$, it can be proved that the grid condenser voltage will closely follow the modulation envelope provided that

$$C_g R_g > \frac{\sqrt{1 - K^2}}{\omega_a K}$$

The time constant $C_g R_g$ may of course be reduced by a reduction of either or both of its factors, but C_g must be large in order to avoid attenuation of the radio-frequency signal, while C_g should be small and R_g large in order to reduce the damping imposed upon the input circuit by the grid leak, and by the input resistance of the valve itself, taking the Miller effect into account. In practice the capacitance of the grid condenser is first decided, say from five to ten times the input capacitance of the valve, e.g. $C_g = 150$ to $300 \mu\mu F$, and the grid leak is then given such a

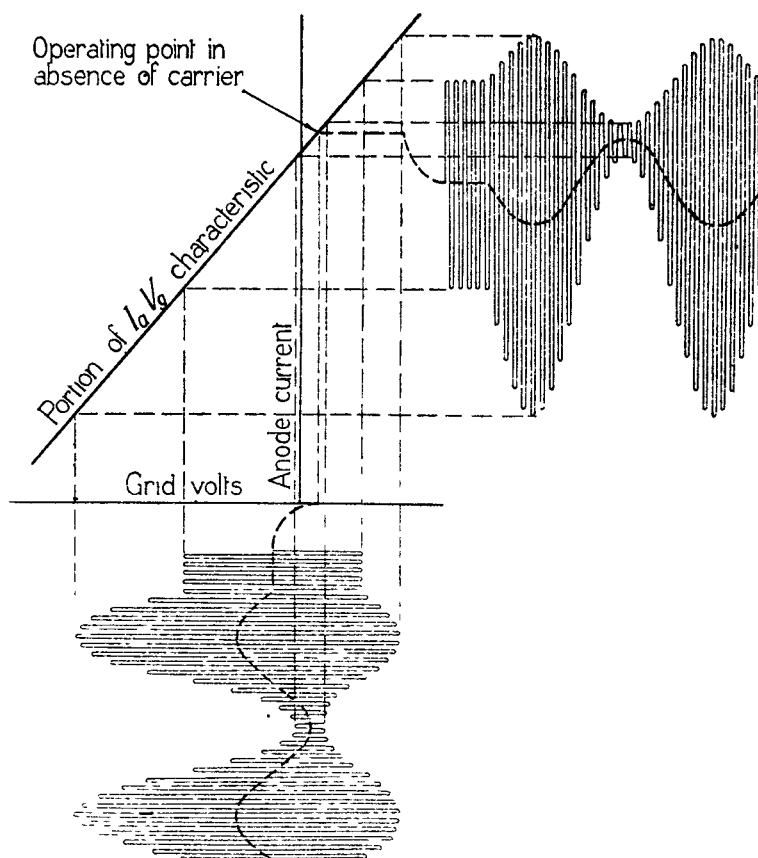


FIG. 36, CHAP. XII —Grid circuit linear rectification; action in anode circuit.

CHAPTER XII.—PARAS. 56-57

value that $C_g R_g$ is of the order of 20 microseconds. The grid leak resistance is therefore of the order of only $\cdot 1$ megohm, which is considerably less than is used for the detection of C.W. or I.C.W. signals. The signal strength is of course somewhat less than with a leak of higher resistance.

56. The maximum permissible input voltage of the grid circuit rectifier is limited by the fact that unless the total anode current variation is confined to the straight portion of the I_a-V_g curve, rectification will take place in the anode circuit and will give rise to distortion. Since the radio-frequency variation is much larger than the audio-frequency variation, it is impossible to obtain a large undistorted audio-frequency power output. The greatest efficiency is achieved by operating with a high H.T. voltage, e.g. 100 to 150 volts, and since the valve must dissipate the whole of the D.C. input during periods of no modulation, a small power valve is preferably used. Even so the H.T. voltage must be considerably less than when the valve is used for power amplification, since under the latter conditions it is operated with considerable negative bias, while as a detector the bias voltage is rarely more than a fraction of a volt negative and may be slightly positive. Where it is possible to provide an adequate input swing to the detector, it is preferable entirely to separate the functions of detection and audio-frequency amplification.

Linear rectification by diode

57. Although for small input voltages the diode acts as a square law rectifier, it was shown in Chapter X that if the input voltage is not allowed to fall below a certain value, depending upon the particular type of diode, the peak anode current is proportional to the anode-filament P.D. Provided therefore that the average depth of modulation is below about 70 per cent.,

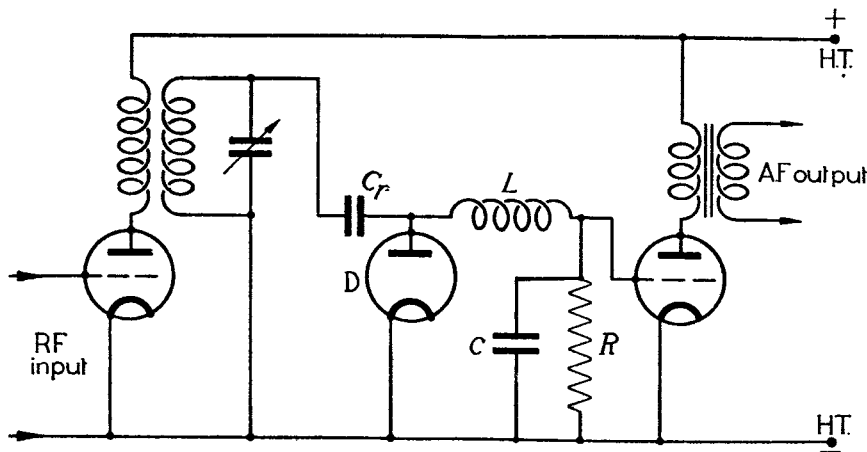


FIG. 37, CHAP. XII.—Diode used as linear rectifier.

and the carrier amplitude is of the order of 10 volts, the rectified output of the diode is practically proportional to the input voltage. In R/T receivers embodying a high degree of radio-frequency amplification, the diode is extensively used as a detector. A typical circuit is given in fig. 37, in which C_r is the reservoir condenser and R the load resistance. The audio-frequency voltages set up across the latter are applied to the grid and filament of the triode which acts as an audio-frequency amplifier. The radio-frequency choke L and condenser C are inserted in order to reduce the amplitude of the radio-frequency voltages applied to the triode. The action of the diode rectifier and the following audio-frequency stage are therefore exactly the same as that of the linear grid rectifier, except that the grid swing on the triode valve is greatly reduced owing to the separation of radio-frequency and audio-frequency functions.

Necessity for radio-frequency gain control

58. It has been shown that in order to reduce distortion to a minimum the input to the detector must lie between certain limits. The radio-frequency amplification required to give this detector input depends upon the minimum field strength upon which the receiver is called upon to operate. The interference level on the receiving site must be taken into account in this respect. For example, in broadcast reception in an industrial centre where the noise level due to electrical devices is very high, programmes of entertainment value can only be received from transmitters giving a field strength of about 10 milli-volts per metre, while in rural areas where electrical interference is negligible a higher degree of amplification can be used and equally good reception obtained from transmitters giving a field strength of only about 0.5 milli-volts per metre. Service R/T communication must be performed with transmitters giving a much weaker field than this, in the presence of an interference level comparable with that of an industrial area, in addition to a high level of local (i.e. non-electrical) noise. In cases where the field strength is so low that the receiver must be operated at its maximum sensitivity, a considerable noise level must be accepted, whereas if the initial field strength is high a reduction in overall gain will effect a considerable improvement in the signal-noise ratio. This constitutes one reason for the incorporation of a gain control in the pre-detector stages. The second object of this control is to avoid overloading the detector valve. To take a numerical example, suppose the detector to operate linearly over the range 0.5 to 5 volts and the receiver to be used on an aerial having an effective height, h , of 5 metres and a magnification χ , between aerial proper and the grid of the first valve, of 5. If the sensitivity of the receiver, i.e. the minimum field strength, $\hat{I}$, which will give 0.5 volt input to the detector, is 4 micro-volts per metre, the input voltage to the first valve must be $\chi h \hat{I}$ volts, or $5 \times 5 \times 4 \times 10^{-6}$ volts. The amplifier is therefore required to give a gain of 5,000, which is easily attained by two tuned stages using screen-grid valves or radio-frequency pentodes. The receiver will only rarely be required to operate at its maximum sensitivity, but may often be called upon to deal with field strengths as high, say 5 milli-volts per metre. The maximum permissible detector input, 5 volts, can then be obtained with a voltage gain of 40, and if this is exceeded the detector will be overloaded. Hence it is necessary to provide a smooth control of the voltage gain between these limits.

Methods of gain control

59. (i) When the radio-frequency stages embody screen-grid valves or radio-frequency pentodes, a certain amount of control may be obtained by variation of either the grid bias voltage or the screen potential. The first method operates by virtue of the fact that the curvature of the $I_a - V_g$ characteristic is considerably greater than that of the triode, i.e. the mutual conductance g_m varies with the mean grid bias, as shown in fig. 38, which is the $g_m - V_g$ curve of a typical radio-frequency pentode. Since for a given dynamic resistance R_d the voltage gain is approximately equal to $R_d g_m$, it is easy to find the variation in gain for various bias voltages, e.g. if the anode load is 50,000 ohms, the gain varies from 80, for values of bias in the region of -0.5 volt, to about 25 when the bias is -3 volts. At the latter point, however, the mean operating point is situated very low down upon the $I_a - V_g$ curve where the curvature is very pronounced, the curvature being proportional to the slope of the $g_m - V_g$ curve. The objection to this method of control is now easily seen. The larger the input voltage is, the greater is the necessity for a long straight portion of the $I_a - V_g$ curve, whereas an increase of negative bias shifts the operating point into a region where the curvature is more pronounced. The result is that although the amplification is reduced, the envelope of the anode current variation is considerably distorted. In addition, cross-modulation will be introduced if an interfering signal is also present. The latter phenomenon will be dealt with later.

(ii) Control of amplification by variation of screen potential suffers from similar disadvantages. A reduction of screen potential reduces the slope of the $I_a - V_g$ curve and therefore reduces the amplification, but the grid base line available for the input voltage swing is reduced in a corresponding degree, the curve moving to the right with a decrease of screen voltage.

CHAPTER XII.—PARAS. 60-61

Thus the operating point is shifted into the curved portion of the characteristic as before. In addition, if the input swing is increased without a corresponding increase of negative grid bias, grid current may flow during a portion of each positive half-cycle.

60. With either of the above methods, it is generally found that operation of the gain control affects the tuning of the receiver to some extent. This raises an important point in practical manipulation. The standard practice is to tune aircraft receivers to the desired frequency on the ground, by means of a suitable heterodyne wavemeter. If the latter is placed very near the receiver, the operator will first pick up the emission (modulated C.W.) with the gain control at maximum sensitivity, and may then proceed to reduce the gain and trim up the tuning simultaneously. This procedure is incorrect, for on returning the gain control to the position of maximum sensitivity the receiver will be slightly de-tuned, and may be incapable of dealing with a very weak signal. The correct method is to pick up the signal with maximum gain, and progressively to move the wavemeter away from the receiver while trimming up,

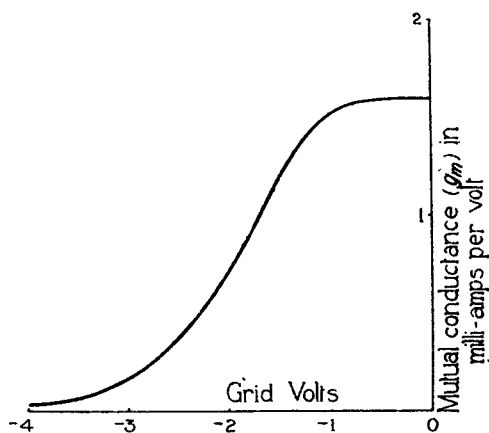


FIG. 38, CHAP. XII.— $g_m - V_g$ curve of R.F. pentode.

operating always with maximum gain. The receiver will then be taken into the air fully capable of dealing with the weakest signals, and will be slightly de-tuned only when very strong signals are being handled.

Cross-modulation

61. Cross-modulation is the name given to a form of interference peculiar to radio-telephony, and will be described with reference to a receiver using screen-grid valves in the radio-frequency stages. If the receiver is tuned to a station giving a high carrier level, and the gain reduced by negative grid bias, the operating point is situated in the curved region of the $I_a - V_g$ characteristic. Any other transmitter which is operating on an adjacent frequency channel may give an appreciable input to the first valve, and the instantaneous grid swing on the latter is then equal to the sum of the two swings. Owing to the anode bend rectification which occurs, the output at the desired carrier frequency will possess side-bands corresponding to the modulation of both signals, and no amount of selectivity in the succeeding circuits will discriminate against this spurious modulation. The effect can therefore only be avoided by a considerable degree of selectivity between aerial and first valve, or by operating in the region of zero bias and controlling the gain by loose input coupling, "lossing" and screen potential variation. One drastic expedient which is sometimes adopted is to fit a switch which disconnects the aerial entirely, leaving the receiver to operate upon the "pick-up" of the coils and wiring.

62. (i) The manner in which cross-modulation is introduced may be shown by assuming the amplifier valve to possess a characteristic represented by an equation of the form

$$i_a = A + B v_g + C v_g^2 + D v_g^3 \dots \dots \dots (1)$$

Let $v_g = W \sin \omega_1 t (1 + K \sin \alpha t) + U \sin \omega_2 t (1 + M \sin \beta t)$,

where W is the amplitude of the "wanted", U the amplitude of the "unwanted" signal, while ω_1 and ω_2 correspond to radio frequencies which are modulated at frequencies $\frac{\alpha}{2\pi}$ and $\frac{\beta}{2\pi}$ to depths of K and M respectively. The important term in introducing cross-modulation is $D v_g^3$.

$$\begin{aligned} v_g^3 &= \{W \sin \omega_1 t (1 + K \sin \alpha t) + U \sin \omega_2 t (1 + M \sin \beta t)\}^3 \dots \dots \dots (2) \\ &= (a + b)^3 \\ &= a^3 + 3a^2b + 3ab^2 + b^3. \end{aligned}$$

Expanding the term corresponding to $3ab^2$ we obtain a component of v_g^3 , say v'_g , where

$$v'_g = 3W \sin \omega_1 t (1 + K \sin \alpha t) \{U^2 \sin^2 \omega_2 t (1 + M \sin \beta t)^2\} \dots \dots \dots (3)$$

Since $U^2 \sin^2 \omega_2 t = \frac{U^2}{2} (1 - \cos 2\omega_2 t)$ and $(1 + M \sin \beta t)^2 = 1 + 2M \sin \beta t + M^2 \sin^2 \beta t$, there will be a component of v'_g , say v''_g , where

$$\begin{aligned} v''_g &= 3W \sin \omega_1 t \times \frac{U^2}{2} \times 2M \sin \beta t \\ &= 3WU^2M \sin \omega_1 t \sin \beta t \\ &= \frac{3WU^2M}{2} \left\{ \cos (\omega_1 - \beta)t - \cos (\omega_1 + \beta)t \right\}, \dots \dots \dots (4) \end{aligned}$$

which represents side-band components associated with the frequency $\frac{\omega_1}{2\pi}$ of the desired signal, but corresponding to the modulation carried by the undesired signal.

(ii) The degree to which cross-modulation is introduced is therefore proportional to the square of the amplitude of the unwanted signal at the grid of the valve and is also proportional to the coefficient of V_g^3 , i.e. to D , in equation (1). It follows that the degree of cross-modulation is approximately proportional to the rate of change of curvature of the valve characteristic. This accounts for the increase of cross-modulation which occurs when the operating point is brought on to a curved portion of the characteristic by the application of large negative bias. Cross-modulation is almost entirely eliminated by the use of variable-mu valves in the radio-frequency stages.

Modulation rise

63. This is another phenomenon which occurs when operating upon a markedly curved portion of the characteristic. Under these conditions, the amplification depends to some extent upon the grid swing, being greater for large inputs than for small. It follows that when a modulated voltage is received, the peaks of the modulation envelope receive greater amplification than the troughs, and in the output the ratio of peak to trough is greater than in the input. This effect varies approximately with the square of the applied carrier amplitude, and is also proportional to the curvature of the $g_m - V_g$ curve. It is therefore of much greater importance in ordinary screen-grid valves and R.F. pentodes than in valves of the type described in the following paragraph.

The variable mutual conductance valve

64. This valve is generally referred to as the variable-mu valve. Its object is to give a considerable control of amplification by grid bias variation, without the attendant disadvantages. Both tetrodes and pentodes may be given the desired characteristics, which are obtained by

CHAPTER XII.—PARAS. 65-66

special design of the control grid. Whereas in ordinary valves the latter is of uniform spacing throughout its length, in the variable- μ valve it is made in two or more sections of varying degrees of spacing. Fig. 39 shows the electrodes of a variable- μ screen-grid valve in cross-section, and it can be seen that the middle portion of the control grid is of comparatively open mesh. It will therefore exercise a smaller degree of control than the upper and lower portions, which are of fairly close mesh. Even if the grid is given a considerably negative potential, therefore, an appreciable anode current will flow, and the $I_a - V_g$ curve tails off slowly with increasing negative bias, instead of possessing a comparatively sharp cut-off. In fig. 40 the $I_a - V_g$ curves of a variable- μ screen-grid valve (full line) and an ordinary screen-grid valve (dotted line) may be compared. The latter can only accommodate, without curvature distortion, a grid swing of about 1.5 volt; to do this the grid bias has a fairly critical value of about -0.75 volt. When given a negative bias of about 15 volts, however, the variable- μ valve will accommodate a grid swing of about 10 volts, the characteristic being practically straight over this range. As the grid bias decreases, the length of the straight portion available for distortionless amplification also decreases. For an input swing of 1.5 volts, the bias may be only about -0.75 volt, as in the screen-grid valve.

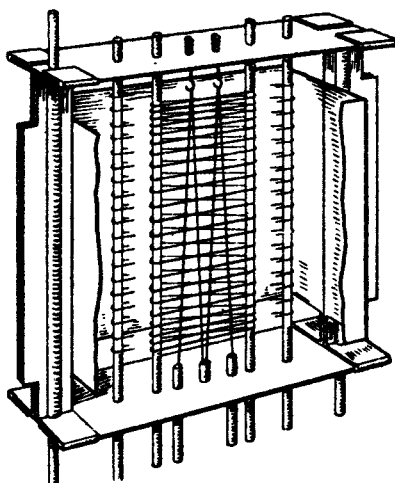


FIG. 39, CHAP. XII.—Electrodes of variable- μ screen grid valve

65. The small inset diagram in fig. 40 shows the variation of mutual conductance with grid bias and may be compared with the $g_m - V_g$ curve of a screen-grid valve (fig. 38). The mutual conductance is seen to be, approximately, inversely proportional to the bias voltage. This is the feature which makes the valve so useful for gain control, for if the bias voltage is varied in such a manner that it is always proportional to the input voltage, the output voltage will remain constant. In modern R/T receivers, therefore, the magnitude of the bias voltage in the radio-frequency stages is often made to depend upon the amplitude of the received carrier. Such a receiver is said to be fitted with automatic volume control (A.V.C.).

Automatic volume control

66. In receivers fitted with automatic volume control, a portion of the carrier frequency input to the detector (or second detector in a super-heterodyne receiver) is rectified, and the resulting current passed through a resistance which is so arranged that an increase in rectified current, i.e. an increase in carrier amplitude, applies a corresponding negative bias to the gain-controlled valve or valves. Considerable radio-frequency amplification must be available, in order that the rectified current in the resistance will be sufficient to provide the maximum bias voltage called for. For this reason, A.V.C. is rarely found in other than super-heterodyne receivers.

For simplicity, however, the principle is illustrated in fig. 41 as it would be applied in a simple receiver comprising one radio-frequency amplifying stage and a linear grid rectifier. The first valve (V.S.G.) is a variable-mu S.G. valve and the second (T) a triode. The latter is the signal rectifier; its input voltage is also applied to the A.V.C. amplifier valve (S.G.) the output of which is applied to the diode (D). This operates as a half-wave rectifier and establishes a P.D. across the reservoir condenser C_1 and resistance R_1 , the upper plate of the condenser being at negative potential with respect to the filament line (L.T. negative).

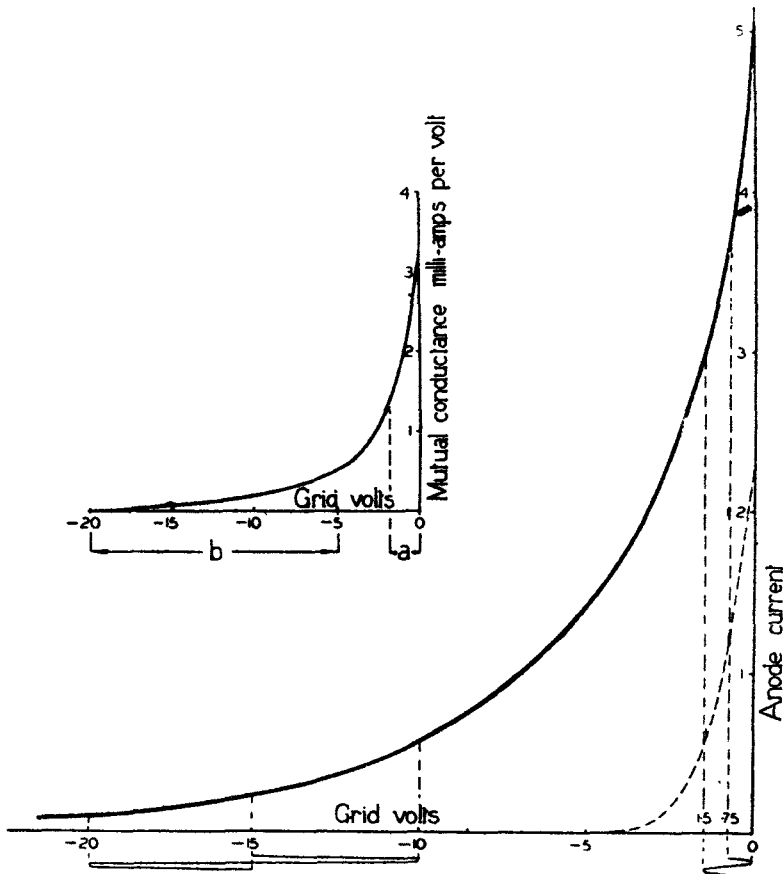


FIG. 40, CHAP. XII.—Characteristic curve of variable-mu screen-grid valve.

This P.D. is applied to the grid of the first valve via the resistance R_p , which is therefore biased to an extent depending upon the amplitude of the input to the diode. The desired effect would not be completely attained by the system so far described, because the A.V.C. would operate to some extent upon all signals, instead of only upon those exceeding a certain amplitude. This can be overcome by applying a suitable bias to the anode of the diode D, as indicated by the inclusion of the battery E in the diagram. In practice it is not convenient to insert a battery at this point, and the necessary delay voltage, as it is called, is obtained by means of a tapping on a resistance connected between the negative L.T. and negative H.T. terminals. (See paragraph 77).

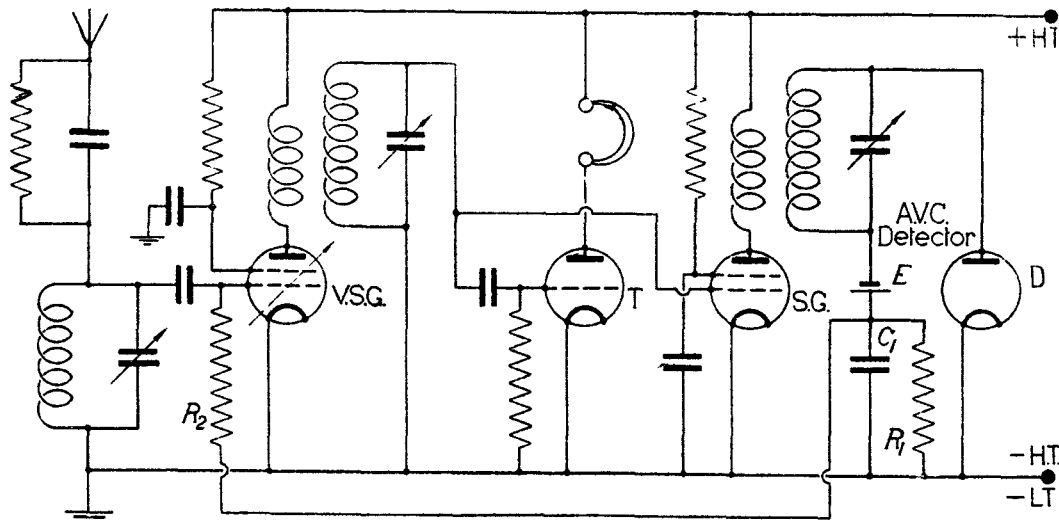


FIG. 41, CHAP. XII.—R/T receiver with A.V.C.

Audio-frequency amplification

67. After detection, it is generally desirable to utilize one or more stages of amplification, the final valve operating as a power amplifier and supplying the reproducing instrument—telephone receivers or loud speaker as the case may be. For voltage amplification, resistance-capacitance coupling is preferred for reasons stated in Chapter XI. An electrical output of the order of 150 milliwatts from the final power amplifier is sufficient to give good speech in two pairs of telephone receivers connected in parallel. This may be obtained from a small power valve, i.e. one capable of dissipating from 500 to 1,000 milliwatts, with an input swing of about 10 volts. A small loud speaker may also be operated in this way, but a high sound level cannot be expected. It is sometimes necessary to obtain a greater power output than can be given by an ordinary power valve. When this is so, several possibilities present themselves. They are the employment of

- (i) A super-power output triode, i.e. a triode capable of dissipating one or more watts.
- (ii) Two or more ordinary power valves in parallel.
- (iii) A pentode.
- (iv) Power triodes or pentodes in push-pull.

The super-power triode

68. This valve differs from the ordinary output triode in that it is designed to accept a greater grid swing. Its anode A.C. resistance is usually low, from 1,000 to 2,000 ohms, and its mutual conductance normal, e.g. about 2.5 milliamperes per volt. The $I_a - V_a$ characteristics of a typical valve of this kind are given in fig. 42. This valve has an anode A.C. resistance of about 1,600 ohms and its mutual conductance is 3 milliamperes per volt, the amplification factor being 5. In the diagram the load line is that of a dynamic resistance of 3,475 ohms, which is approximately twice the anode A.C. resistance. The permissible dissipation is 3 watts and the operating potentials are $E_a = 145$ volts, $E_g = -17.5$ volts. Without entering the regions of anode current curvature and grid current flow, the permissible grid swing is 35 volts, and the maximum undistorted output, as shown by the shaded area, is 485 milliwatts. The efficiency is therefore only 16 per cent. For comparison the area corresponding to the output obtained with a grid swing of 11 volts is also shown in vertical shading; this is equal to 50

milliwatts. A small power valve with a grid swing of this order will give an output of about 100 to 150 milliwatts, thus, unless the super-power valve is supplied with an ample input voltage, the power output obtainable may be less than that given by a small power valve. The figure of merit (i.e. $\frac{\mu^2}{r_a}$) for the valve under discussion, is only $\cdot 015$ against $\cdot 038$ for the valve, V.R.22.

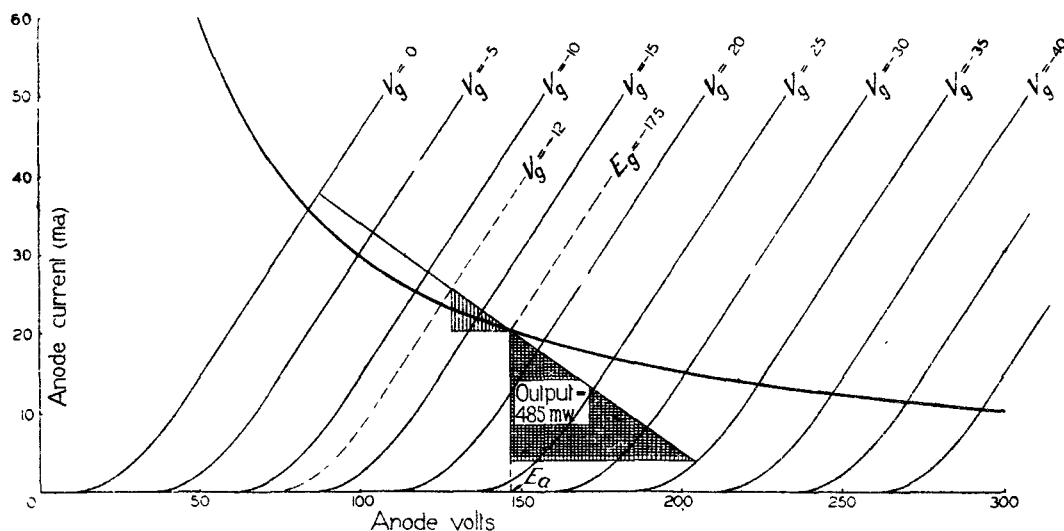


FIG. 42, CHAP. XII.— $I_a - V_a$ curves of super-power triode.

Two power valves in parallel

69. (i) With this arrangement the amplification factor is unchanged, but the effective anode A.C. resistance is only one-half that of a single valve; it must be understood that only valves of the same type may be connected in parallel. The optimum load for maximum undistorted output is equal to r_a and the maximum grid swing that of a single valve. The anode current consumption will be doubled, as will also the power taken from H.T. supply, hence this method is unsuitable for use where only dry cell or inert batteries are available. The power output is twice that obtainable from one valve of the same type, if the optimum loading conditions are achieved in each case.

(ii) Provided that the required grid swing is available, therefore, it is preferable to use a super-power valve rather than two power valves in parallel. Both arrangements suffer from the following disadvantages :—

(a) A large steady anode current flows in the output choke or primary winding of output transformer, and unless specially designed the core may be magnetically saturated. Even if this is not so, the high flux density will cause heavy hysteresis loss; this loss varies at different instants during each cycle, and thus gives rise to amplitude distortion, while the reduction of inductance due to the fall of incremental permeability with increase of flux density causes additional frequency distortion.

(b) Ordinary H.T. batteries are incapable of supplying this anode current. Either super-capacity batteries, accumulator batteries, or mains supply may be used.

(c) A large alternating component of anode current flows through the H.T. battery (or other supply device). Earlier stages of the receiver must therefore be very thoroughly decoupled, otherwise low-frequency oscillations may be established. It is sometimes possible

CHAPTER XII.—PARA. 70

to prevent this by reversing the connections to one winding of an intervalve or output transformer, but this is not the best practice, for the energy transfer which previously maintained the oscillation will then impose considerable damping and consequent reduction in overall amplification.

Use of pentode valve

70. The pentode is a five electrode valve, and its characteristics have already been discussed in Chapter VIII. It is chiefly used when greater power output is required than is obtainable from a power valve for the same input grid swing, without serious increase of either H.T. voltage or anode current. Considerable care is necessary in the choice of correct load impedance. Fig. 43 shows the $I_a - V_a$ curves of a pentode, with load lines AB, CD, EF, representing anode circuit resistance loads of approximately 40,000 ohms, 13,700 and 4,000 ohms respectively. The maximum permissible input voltage with this particular valve has been taken as 5 volts (peak), and the working H.T. voltage as 150 volts, the anode current in the absence of an applied input voltage being 10·5 milliamperes. Taking the 40,000 ohm load, an input of 5 volts (peak)

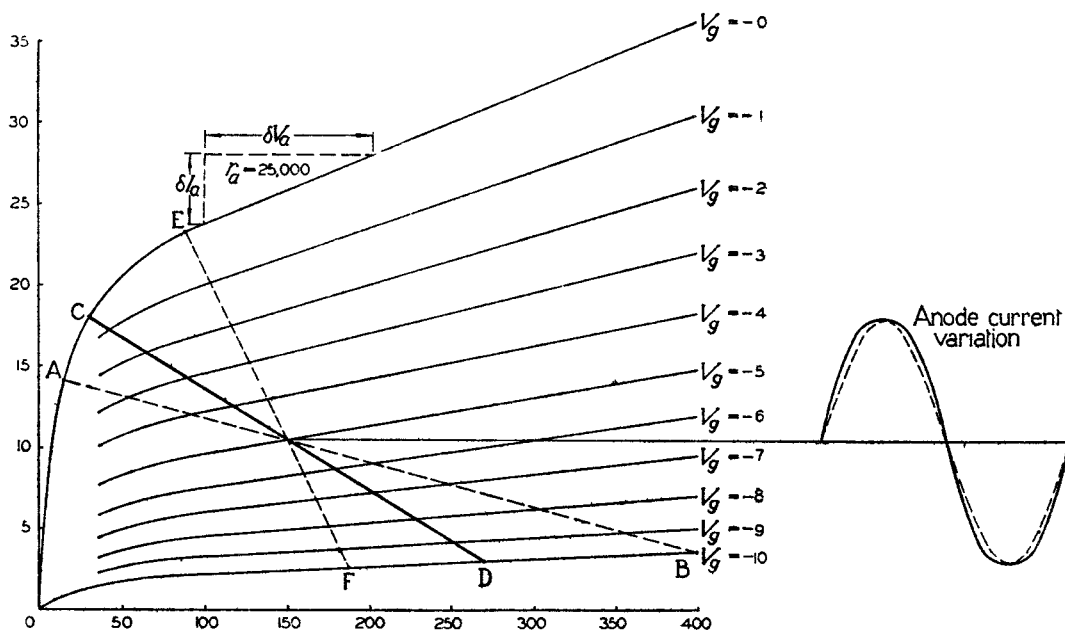


FIG. 43, CHAP. XII.— $I_a - V_a$ curves of output pentode.

will cause the following variations of anode current and anode-filament P.D., namely (i) at the positive peak of input voltage the anode current will rise to 13·7 milliamperes and the anode-filament potential will fall to 12 volts, (ii) at the negative peak of input voltage the anode current will fall to 5 milliamperes and the anode-filament P.D. will rise to about 400 volts. As the variations of current and voltage on the positive and negative half-cycles are quite dissimilar, distortion will obviously occur. With an anode circuit load of this magnitude, another undesirable effect will also arise; when the anode-filament P.D. falls as low as 12 volts, the flow of electrons between filament and anode will be retarded and a negative space charge will form in the vicinity of the anti-secondary or suppressor grid; the control grid then ceases to effect any control upon the anode current. The combination of these two effects will inevitably give rise to severe distortion.

71. (i) If the load resistance is too low, the variations of anode current and anode-filament P.D. during the positive and negative half-cycles, e.g. on the load line EF, will again be unsymmetrical and distortion will occur. With a suitably chosen resistance, however, the excursions of both anode current and anode-filament P.D. will be nearly symmetrical about the mean operating point. Thus with the 13,700 ohm load, the anode current increases and decreases by 8 milliamperes as the grid voltage undergoes a complete cycle, the variation of anode-filament P.D. being also symmetrical, and the output will be comparatively free from distortion. It will also be observed that the minimum anode-filament P.D. is about 35 volts, which is sufficiently high to prevent the negative space charge effect at the anti-secondary grid.

(ii) The anode A.C. resistance of a pentode varies greatly with grid potential; if specified, it is usually taken as the ratio $\frac{dV_a}{dI_a}$ measured at the operating H.T. voltage on the curve corresponding to $V_g = 0$. In the diagram the value of r_a at this point is approximately 25,000 ohms, i.e. twice the value of load resistance giving maximum undistorted output. In general, the optimum load of the pentode may be taken as one-half the anode A.C. resistance as defined above, or alternatively as being equal to the D.C. resistance of the valve at the operating point. In fig. 43 this is $\frac{150 \text{ volts}}{10.5 \text{ milliamperes}} = 14,300 \text{ ohms}$. In this respect there is a marked difference between optimum conditions for pentode and triode valves, and one cannot be substituted for the other without complete re-design of the output stage.

(iii) On the right of fig. 43 is shown, in solid line, the anode current variation caused by a sinusoidal variation of grid voltage when operating with the 13,700 ohm load, together with a sine curve (dotted line) for comparison. Since the wave-form differs from the sine curve in the same manner in both half-cycles, odd harmonics must be present; as a rule the third harmonic is most conspicuous. For the same percentage distortion, this is more objectionable than the second harmonic, which is exactly one octave higher than the fundamental. The small scale of the drawing does not admit of very accurate computation, but assuming that only the third harmonic is present, the distortion is about 7 per cent.

Valves in push-pull

72. (i) The use of valves in push-pull connection has already been considered in Chapter IX. When used for audio-frequency amplification, the circuit is as shown in fig. 44. The output impedance of the previous stage is the primary of a special form of iron-core transformer having a centre-tapped secondary winding, and one-half of the secondary voltage is applied to grid and

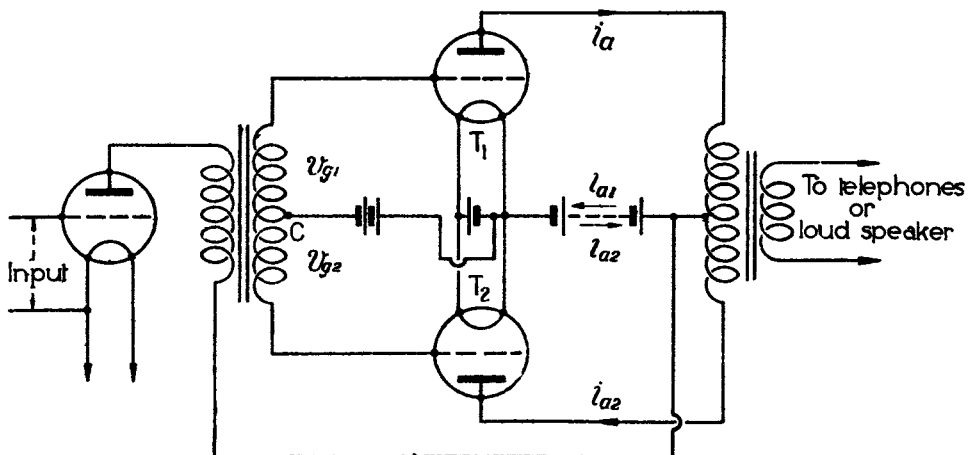


FIG. 44, CHAP. XII.—Push-pull output.

CHAPTER XII.—PARAS. 73-74

filament of each valve. In the anode circuit of each valve is connected one half of the primary winding of the output transformer, the H.T. supply being fed to both valves via the centre tap. Either Class A or Class B operation may be used. Referring to fig. 44, suppose an alternating voltage to be induced in the secondary winding of the input transformer. Whenever the grid of the valve T_1 is at a positive potential with respect to the centre point C of the winding, the grid of the valve T_2 will be at an equal negative potential, and vice versa. In the absence of a signal voltage, a steady electron current flows from filament to anode of each valve, its value depending upon the H.T. supply voltage and grid bias. Provided the two valves have identical characteristics, these steady anode currents will be equal, and will flow in opposite directions round the core of the output transformer, hence no resultant magnetic flux will be established in the latter. This in itself will remove one possible cause of distortion.

(ii) When a signal voltage is applied to the primary of the input transformer, equal voltages v_{g1} , v_{g2} will be applied to the grids of the two valves, in antiphase; the resulting anode currents are i_{a1} and i_{a2} , and these currents flow in the two halves of the output transformer primary. As an increase in the value of i_{a1} is accompanied by a simultaneous decrease in the value of i_{a2} , and vice versa the induced E.M.F. in the two halves of the winding will be in phase and of equal amplitude, hence the two voltages may be considered to assist each other in producing an alternating flux in the core and a consequent induced secondary E.M.F.; on the other hand the alternating currents i_{a1} and i_{a2} are in opposition so far as the common connection between negative of filament and centre tap of output winding is concerned, and provided these currents are equal no alternating current will flow through the H.T. source.

Class A amplification

73. Assuming that the two valves have identical characteristics, the total anode circuit load, often referred to as the "anode to anode" load, should be rather less than $4r_a$, and the grid bias rather more than would be used with a single valve, because the anode current may be allowed to swing to a somewhat lower value without introducing appreciable distortion. The power output of a perfectly balanced Class A amplifier is twice the output obtainable from a single valve (assuming optimum load and equal grid swing in each case). As the push-pull arrangement will accept a rather greater grid swing, it would appear possible to obtain nearly three times the output given by a single valve, for the same degree of distortion, but this is rarely so in practice owing to the difficulty of obtaining two valves with identical characteristics. The advantages of the Class A push-pull amplifier over the previously discussed alternatives are as follows:—

- (i) No appreciable variation of battery current during each cycle, and consequently little tendency to transfer energy to earlier stages by battery coupling.
- (ii) The absence of a steady magnetizing current in the primary winding of the output transformer, permitting the use of a cheaper and lighter transformer.
- (iii) A slight increase in permissible grid swing, compared with that of a single valve or two valves in parallel. The super-power valve however has the advantage in this respect.

Quiescent push-pull

74. This term is frequently used to denote the type of circuit in which the grids are biased to cut-off point, so that in the absence of an alternating input voltage the anode current is negligible. This is correctly termed Class B amplification, but this appellation is generally applied to a special arrangement which is described later. Either triodes or pentodes may be used in quiescent push-pull. In either case the permissible grid swing is approximately double that of a single valve, and theoretically the output should be quadrupled. Since however each half of the output transformer is energized only during alternate half-cycles, the output power is only slightly greater than that obtainable under Class A conditions. Although the second harmonic variations of anode current are in opposition in the load, they are in phase in that

portion of the anode circuit which is common to both valves, i.e. the H.T. battery, and complete decoupling of earlier stages is therefore essential. The principal disadvantage of quiescent push-pull working is the difficulty of securing a perfectly matched pair of valves.

The Class B valve.

75. The form of quiescent push-pull arrangement generally referred to as "Class B" amplification utilizes a special valve which comprises two carefully matched triodes in a single envelope, the filament being common to both electrode systems. Only a small negative bias, say 1.5 volt, is used, some valves being designed to operate with zero bias, and grid current flows during a considerable portion of each cycle. The input impedance is correspondingly low, and it is necessary to provide a power input rather than merely a wattless input voltage as is generally the case. The preceding stage is operated as a power amplifier and is called the driver of the Class B valve, to which it is coupled by an iron-core transformer with centre-tapped secondary. The transformer must be designed to match the input impedance R_i of the Class B

valve to the A.C. resistance r_a of the driver valve, its ratio being $\sqrt{\frac{R_i}{r_a}}$, while the resistance of its secondary winding must be low compared to R_i . The output transformer must also be carefully designed in order that its internal resistance and leakage reactance may be small, otherwise considerable distortion arises owing to the variation of effective load at different frequencies. The remarks regarding variation of battery current in quiescent push-pull amplifiers are equally applicable to Class B. Quiescent push-pull and Class B amplification are chiefly used where only dry-cell batteries are available for H.T. supply. Any device having considerable internal impedance, e.g. a so-called H.T. eliminator, can only be used if special steps are taken to ensure that the terminal P.D. does not vary with the load current to any appreciable extent.

The paraphase amplifier

76. In the preceding paragraphs it has been assumed that the input circuit of the push-pull amplifier is perfectly symmetrical. It is difficult to ensure this at the higher audio-frequencies, owing to the difference between the capacitance of each end of the transformer winding with respect to earth. The varying iron losses in this transformer also cause amplitude distortion. For both reasons it would appear advantageous to precede the push-pull output stage by a resistance-loaded voltage amplifier. This is achieved in the paraphase amplifier by means of a phase reversing valve, the connections being given in fig. 45. Here T_1 is the amplifying valve following the rectifier; the resistance R_1 , grid condenser C_1 , and resistance R_2 forming an output network. A portion of the voltage across R_2 is applied to the grid and filament of the phase reversing valve T_2 , which has the output network R_3 , C_2 , R_4 . When the two networks are correctly balanced, equal voltages are developed across the resistances R_2 and R_4 , in antiphase, and these are applied to the push-pull (Class A) triodes T_3 , T_4 . The correct balance is obtained by adjusting the tapping point on the resistance R_2 . To do this a sinusoidal input is supplied to the valve T_1 and the operator listens in the telephone receivers which are connected in the common anode lead of the output stage. When a correct balance is obtained there will be no current variation in this circuit and consequently no sound output from the receivers.

77. The circuit diagram of a typical super-heterodyne R/T receiver embodying several of the features discussed in previous paragraphs, is given in fig. 46. An aperiodic aerial coupling supplies the input circuit L_1 , C_1 of the frequency changer which is a triode-pentode valve, the pentode section having variable-mu characteristics. The triode section of the frequency changer acts as the oscillator valve in conjunction with the circuit L_4 , C_4 and reaction coil L_5 . The band-pass filter L_2 , C_2 , L_3 , C_3 is tuned to the intermediate frequency and is not adjustable. The intermediate frequency amplifier valve is a variable-mu radio-frequency pentode. Its output circuit supplies the double-diode detector valve. The anode A_1 rectifies the signal voltage, the output voltage being developed in the circuit R_1 , C_5 , R_3 . The latter resistance is a kind of potentiometer, and is used as a control of the average output sound level, this level being then

CHAPTER XII.—PARA. 78

maintained automatically. The automatic volume control operates as follows : rectification takes place at the anode A_2 of the second detector, and a voltage proportional to the carrier amplitude is developed across the resistance R_3 . This voltage is applied, via the resistances R_4 , R_5 to the control grids of both frequency changer and intermediate frequency amplifier valves. The

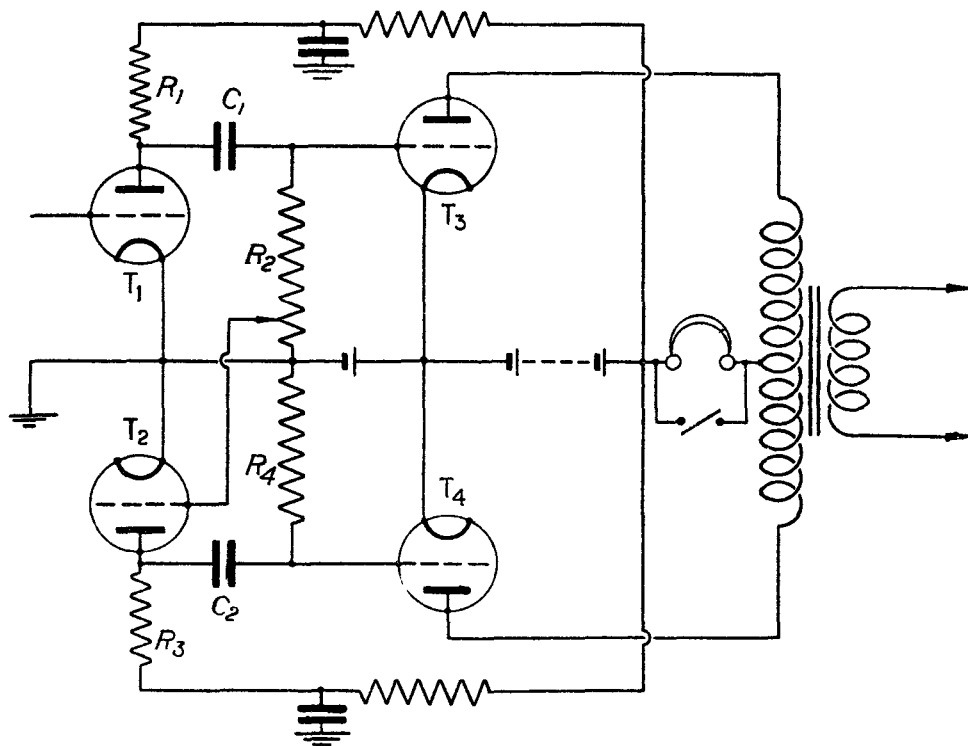


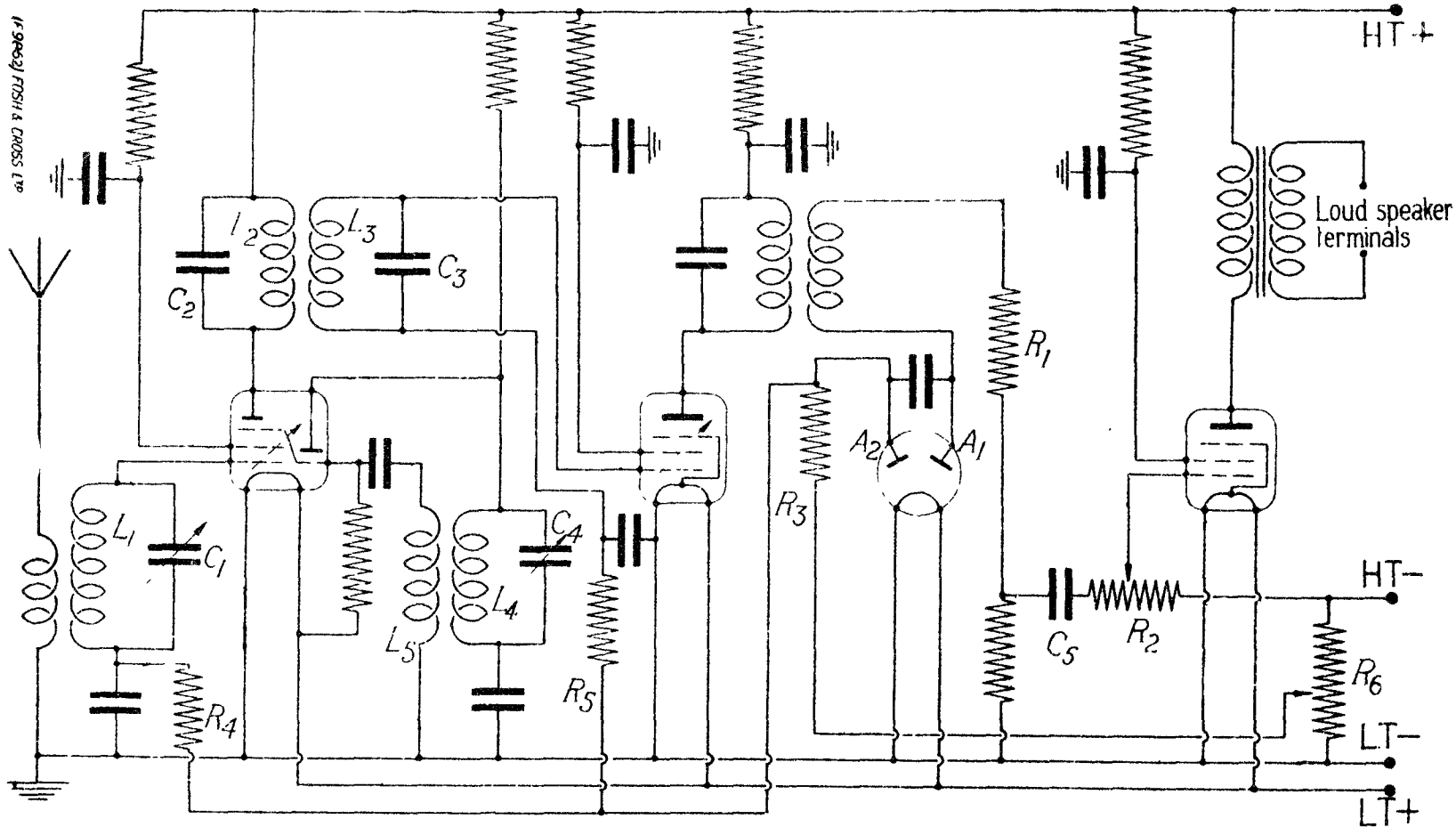
FIG. 45, CHAP. XII.—Paraphase amplifier.

anode A_2 of the second detector is maintained at a mean negative potential with respect to the filament by means of a tapping on the resistance R_5 thus giving a definite delay to the operation of the A.V.C. This resistance also serves to give the required bias to the pentode output valve.

PHASE AND FREQUENCY MODULATION

Phase modulation

78. Phase modulation is performed by maintaining the wave at a constant amplitude, while introducing a cyclical phase shift with reference to the phase of the carrier under non-modulated conditions, the phase shift being proportional to the amplitude of the modulating current or voltage. Such a wave is represented by the heavy line of fig. 47, while the relative phase of the unmodulated wave, which is assumed to be of sinusoidal wave-form, is also shown in the diagram. In the particular conditions illustrated the modulated wave advances in phase during the first quarter of a cycle of the modulating frequency, and then commences to fall back, momentarily assuming its original phase at the end of the first half-cycle. During the next half-cycle the modulated wave lags behind its normal phase, the lag reaching a maximum value at the negative peak of the audio-frequency cycle, after which the phase angle decreases.



SUPER-HETERODYNE RECEIVER FOR R/T

At the end of one complete cycle of modulation the wave has the same phase as it would possess in the absence of modulation. If the phase shift introduced by the modulation at any instant is $\varphi(t)$ radians, a phase-modulated current can be represented by the equation

$$i = \mathcal{P} \sin \{ \omega_r t + \varphi(t) \}$$

$\frac{\omega_r}{2\pi}$ being the carrier or unmodulated frequency. The angle $\varphi(t)$ is not constant but varies sinusoidally at the audio-frequency, $\frac{\omega_a}{2\pi}$. If we define the modulation ratio M in such a manner

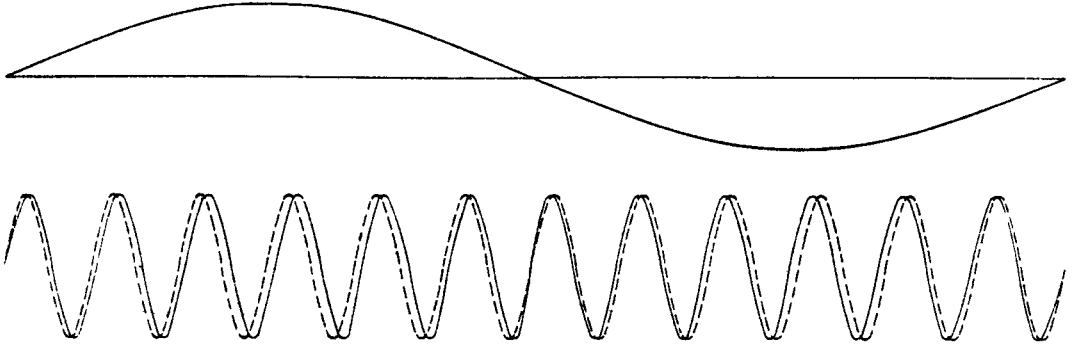


FIG. 47, CHAP. XII.—Phase-modulated wave.

that a modulation ratio of unity will cause a peak phase shift of one complete radio-frequency cycle or 2π radians, $\varphi(t) = 2\pi M \sin \omega_a t$ and therefore

$$i = \mathcal{P} \sin \{ \omega_r t + 2\pi M \sin \omega_a t \}$$

Phase modulation is rarely if ever deliberately adopted for communication purposes, but occurs to some extent in amplitude-modulated systems, as explained in paragraph 40. It is particularly liable to occur when a modulated amplifier is employed, since the magnitude and phase angle of the load impedance generally varies to some extent with the frequency. It is chiefly of importance because of its close resemblance to frequency modulation.

Frequency modulation

79. This type of modulation is achieved by maintaining the constant amplitude of the radiated wave, while varying the frequency in accordance with the amplitude of modulation. Such a wave is shown in fig. 48 which should be compared with the phase-modulated wave of fig. 47. It will be observed that the two waves are very similar in character. The difference between phase and frequency modulation lies partly in the amount of phase shift during any half-cycle of the modulation frequency. If the phase shift in a phase-modulated signal is 2π radians, and there are 1,000 radio-frequency cycles in each half-period of modulation, there will be only 499 radio-frequency cycles in the first quarter period and 501 cycles in the second quarter period of modulation, 501 cycles during the third quarter period and 499 during the fourth, so that the average radio frequency during each half period is equal to the unmodulated or carrier frequency. In a frequency-modulated wave, however, the peak phase shift may be more than 2π radians, and there may be many more radio-frequency cycles during one half period of modulation than in the half period of opposite sign. Hence the average radio frequency during any one half period of modulation is not equal to the frequency of the unmodulated oscillation. It can be shown that the equation representing a frequency-modulated current is

$$i = \mathcal{P} \sin \left\{ \omega_r t + \frac{f_r M}{f_a} \sin \omega_a t \right\}.$$

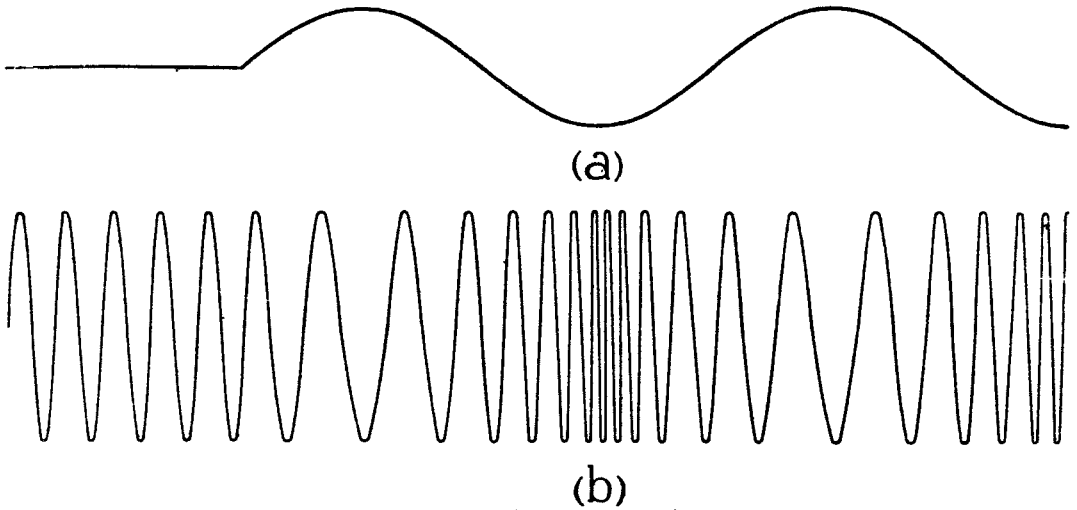


FIG. 48, CHAP. XII.—Frequency-modulated wave.

80. The frequency-modulated wave may therefore be regarded as a particular form of phase-modulated wave, in which the peak phase shift is $\frac{f_r M}{f_a}$ instead of $2\pi M$, i.e. for a given carrier frequency f_r and modulation ratio M the peak phase shift is inversely proportional to

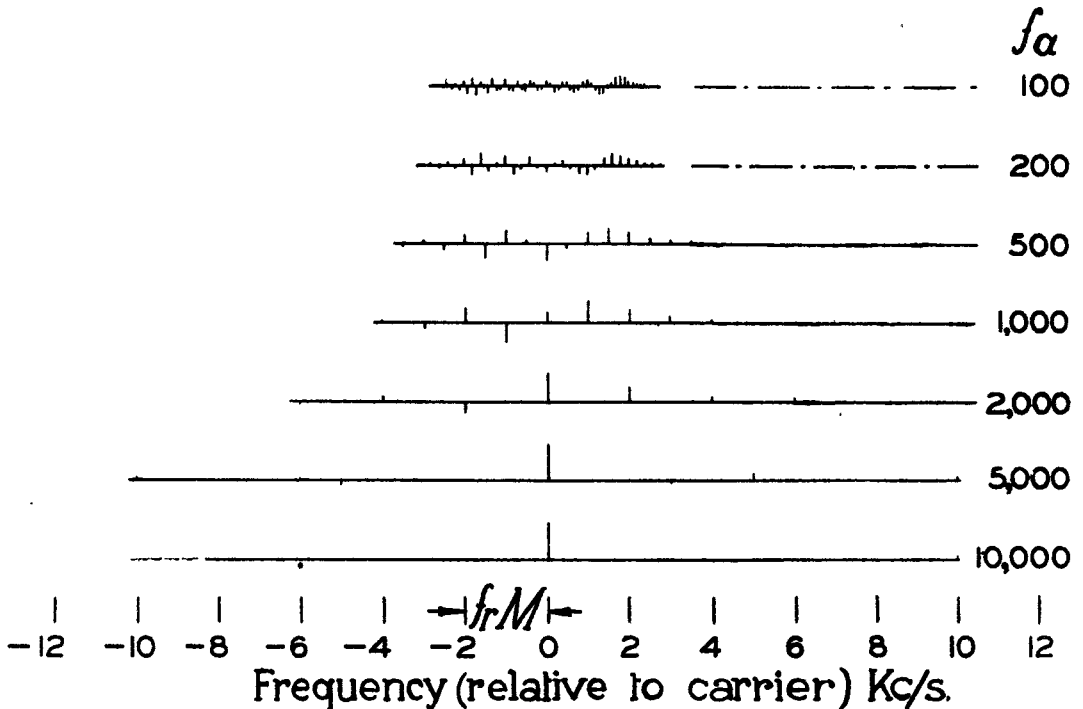


FIG. 49, CHAP. XII.—Amplitude and phase of components of frequency-modulated wave.

the modulation frequency f_m , instead of being a constant for all modulating frequencies as in phase modulation. The expression $f_r M$ will be referred to as the frequency deviation. As in the case of amplitude modulation, the wave may be resolved into the sum of a number of components. The frequencies of these components are f_r , $f_r \pm f_m$, $f_r \pm 2f_m$, $f_r \pm 3f_m$, $f_r \pm 4f_m$, etc., representing a carrier with side-bands. Modulation by a single audio frequency thus produces an infinite number of side-frequencies instead of only one pair. The relative magnitudes and phases of the carrier and side-band components, for a frequency deviation of 2 kc/s and for various modulation frequencies, are shown in fig. 49. When the peak phase shift is less than one radian, the amplitude of the side frequencies $f_r \pm f_m$ is approximately proportional to the frequency deviation $f_r M$, and the carrier component of frequency f_r has an amplitude practically equal to the amplitude of the unmodulated wave. If however the peak phase shift is greater than one radian, the carrier component decreases in amplitude and may be very small or even zero, while the components of frequency $f_r \pm 2f_m$, $f_r \pm 3f_m$, etc. may, be very prominent, and there may be side-bands extending up to the extreme limits of the frequency deviation. For practical purposes, the total band width occupied by a frequency-modulated wave may be taken as rather more than twice the frequency deviation or rather more than twice the modulation frequency, whichever is the greater. It is seen therefore that frequency modulation is no solution to the problem of frequency allocation, the total band width required being somewhat greater than that occupied by an amplitude-modulated transmission of comparable quality.

81. To some extent, frequency modulation must occur in a nominally amplitude-modulated system, particularly in circuits in which the carrier frequency varies with the instantaneous power taken by the oscillatory circuit. The deliberate adoption of this form of modulation for high frequency transmission appears to offer certain advantages, but sufficient experience has not yet been gained to make it clear whether these are offset by its disadvantages. A frequency-modulated system of high quality demands a much more elaborate circuit than an amplitude-modulated system. One of the earliest attempts to realize a frequency-modulated wave consisted of a master-oscillator transmitter in which a condenser microphone of capacitance C_m was connected in parallel with the tuning capacitance C of the master-oscillator, but such a transmitter will produce a purely frequency-modulated wave only (i) if the ratio $\frac{C_m}{C}$ is vanishingly small, so that the frequency deviation and equivalent depth of modulation approach zero, and (ii) if the frequency of the unmodulated oscillation is extremely constant.

82. In a high fidelity frequency-modulated system, then, the following conditions must be satisfied.

- (i) The mean frequency, f_r , must be stable.
- (ii) There must be no amplitude modulation.
- (iii) The frequency deviation must be independent of the modulation frequency and directly proportional to the amplitude of the modulating current or voltage.

Condition (ii) is comparatively easy to satisfy at high frequencies where the ratio of frequency deviation to mean radio frequency can be kept quite small, say one part in one thousand, for the resonance curve of the master-oscillator circuit will be practically flat over this limited deviation. Condition (iii) can be met by a careful consideration of the method of modulation control. Condition (i) is not so easily satisfied; for instance, if an attempt is made directly to vary the frequency of the master-oscillator in the manner outlined in the previous paragraph, it follows that the oscillator must be of low inherent stability.

83. (i) During the last few years, a method has been developed which appears completely to solve this problem. It depends upon the close resemblance between phase and frequency modulation. It has already been stated that in a phase-modulated signal the peak phase shift is $2\pi M$ and in a frequency-modulated signal is $\frac{f_r M}{f_m}$ being therefore inversely proportional to the modulating frequency. In the system to be described, the modulating voltage is caused to shift the phase of a current derived from a source of constant frequency by an amount which

CHAPTER XII.—PARA. 83

is directly proportional to the amplitude of the modulation and inversely proportional to its frequency. For reasons which will be apparent later, the modulated wave is then subjected to considerable frequency multiplication before it is applied to the final power amplifier. The initial process is to produce the required phase shift, the circuit arrangements to this end being shown in fig. 50. The master-oscillator may have a frequency of from 50 to 100 kc/s, and an E.M.F. from this source is supplied to the balanced modulator T_a , T_b , and also, in synphase, to a radio-frequency amplifier valve T_1 . The balanced halves L_a , C_a , L_b , C_b of the output circuit of the modulator are acceptor circuits for the master-oscillator frequency (parallel feed being employed by means of the resistances R_a , R_b) and the anode currents in the coils L_a , L_b are in phase with the corresponding grid-filament input voltages. The output from the modulator feeds a modulation amplifier valve T_2 through an inductive coupling, the effect of the latter being that the output voltage of T_2 is 90° out of phase with the output of the master-oscillator.

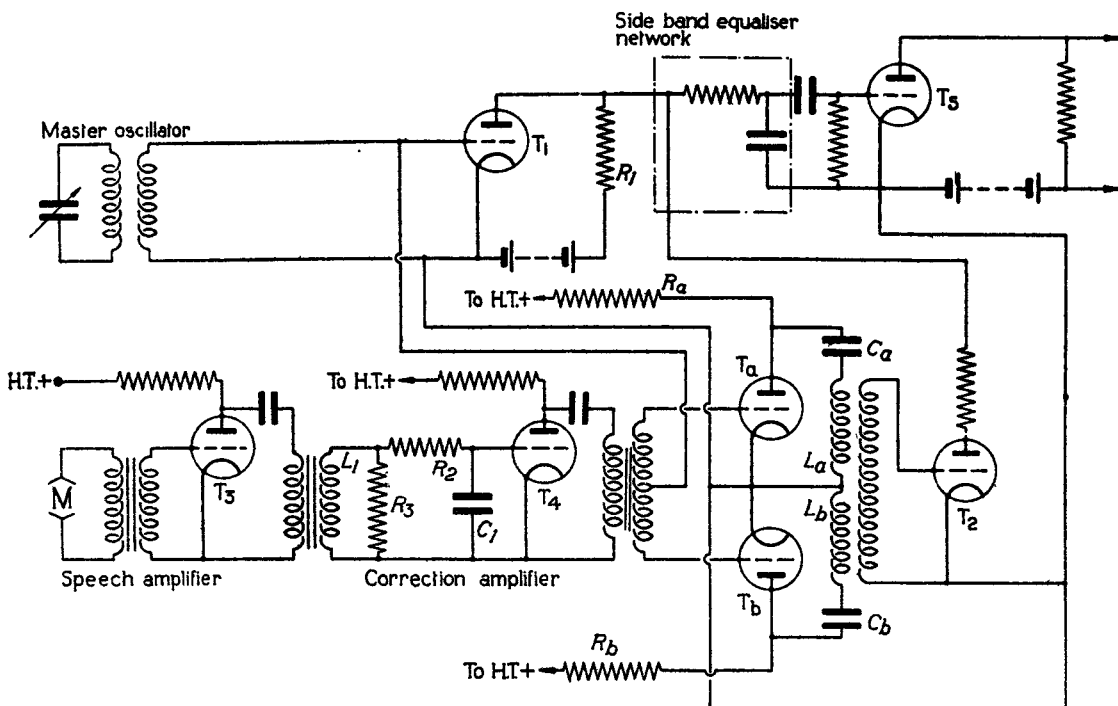
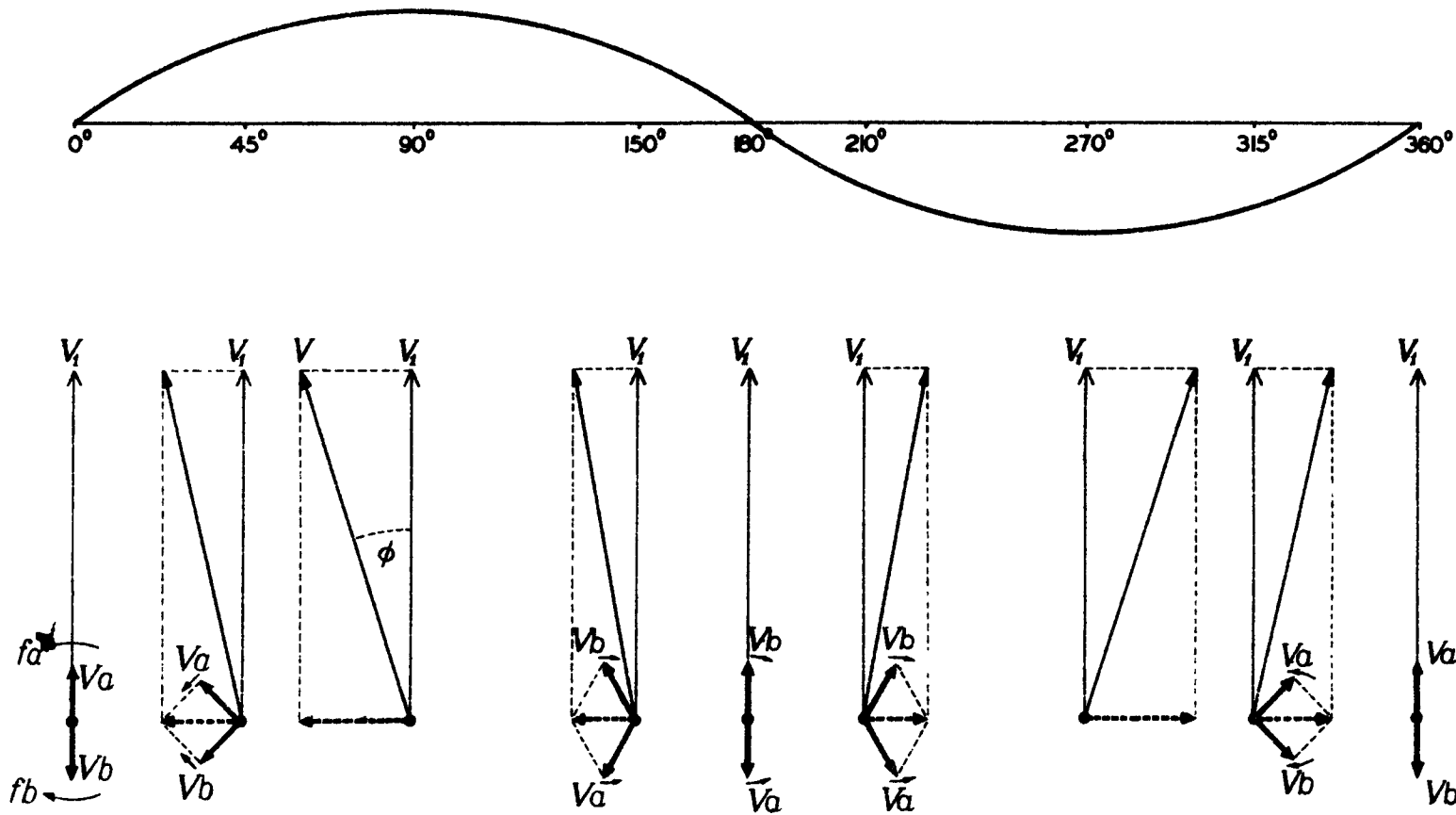


FIG. 50, CHAP. XII.—Modulating stages, frequency-modulated transmitter.

The unmodulated amplifier valve T_1 and modulated amplifier valve T_2 work into a common resistance load R_1 and the voltages developed by the two amplifiers, between the ends of this resistance, are in quadrature. The vector diagrams of fig. 51 illustrate the changes in phase (and incidentally in amplitude) of the P.D. across the resistance R_1 . The vector V_1 represents the voltage due to the valve T_1 while the vectors V_a and V_b represent the output voltages of the valve T_2 , due to the valves T_a , T_b of the modulator. The vector V_1 revolves at ω_r , V_a at $\omega_r + \omega_a$ and V_b at $\omega_r - \omega_a$ radians per second. With respect to the vector V_1 , V_a rotates at $+\omega_a$ radians per second and V_b at $-\omega_a$ radians per second, so that if V_1 is considered to be stationary, V_a and V_b may be considered to rotate f_a times per second in the counter-clockwise and clockwise directions respectively. The peak voltage across the resistance R_1 at several different instants during one audio-frequency cycle is therefore as shown in the diagram. It will be observed that the phase of the resultant voltage varies between ϕ and zero during this period, i.e. ϕ is the maximum phase shift, occurring at the peak of the audio-frequency cycle.



VECTOR DIAGRAM SHOWING PRODUCTION OF PHASE MODULATION

(ii) Provided that φ is not allowed to exceed about 30° , it will be very nearly proportional to the magnitude of the sum of the modulating components. Referring to the diagram, it will be seen that $\tan \varphi = \frac{V_a + V_b}{V_1}$. As φ is made smaller and smaller, so $\tan \varphi \rightarrow \sin \varphi \rightarrow \varphi$, i.e. for small phase shifts $\varphi \doteq \frac{V_a + V_b}{V_1}$. The magnitude of the resultant voltage V varies between the values V_1 and $\sqrt{V_1^2 + (V_a + V_b)^2}$, so that there is a slight degree of amplitude modulation. This again is small if $V_a + V_b$ is small compared to V_1 and in any case is easily removed by a limiting stage.

84. Hitherto we have considered the phase shift to be the same for any value of f_a whereas if the final result is to be truly a frequency-modulated wave, the phase shift must be inversely proportional to f_a . This is achieved by means of the audio-frequency correction network associated with the audio-frequency amplifier valve, T_4 , preceding the balanced modulator (fig. 50). The first valve T_3 serves merely to couple the microphone to the attenuation network $L_1 C_1 R_2 R_3$. The reactance of C_1 at the lowest modulation frequency is small compared to the resistance R_2 . Let a voltage $v = V \sin \omega_a t$ be applied to the series circuit $C_1 R_2$. The P.D. across the condenser will be

$$v_c = \frac{v}{\sqrt{R_2^2 + \left(\frac{1}{\omega_a C_1}\right)^2}} \times \frac{1}{\omega_a C_1}$$

As $\frac{1}{\omega_a C_1}$ is always small compared to R_2 , v_c is approximately equal to $\frac{v}{\omega_a C_1 R_2}$, i.e. $v_c \propto \frac{1}{\omega_a}$, which is what is required. As the maximum phase shift is less than 30° , and this is only permissible on the lowest audio-frequencies, it is necessary to appreciate the order of the phase shift which will be obtained when the highest audio-frequencies are transmitted. If the band to be covered is 100 to 5,000 cycles per second, and the phase shift for 100 cycles per second is 25° , at 5,000 cycles per second it will be $25^\circ \times \frac{100}{5,000} = 0.5^\circ$, and for a wider band it would be still less. This amount of phase shift would produce side bands of negligible amplitude; the apparent difficulty is however removed by frequency multiplication.

85. The conditions for the distortionless amplification of a frequency-modulated wave are

- (i) the gain must be constant over the whole band width occupied by the signal.
- (ii) the phase change introduced by the amplifier must be directly proportional to the frequency.

These conditions ensure that the relative phases and amplitudes of the side bands are unchanged by amplification. It will easily be seen that frequency multiplication increases the frequency deviation in the same ratio as the mean frequency. It may be taken that if the side-bands in the final output are to be of sufficient amplitude to give the equivalent of 100 per cent. amplitude modulation, a phase shift of 45° will be required at the highest audio-frequencies. In the example taken above where the phase shift of the modulator stage is only 0.5° , it would be necessary to employ a frequency multiplication of $\frac{45^\circ}{0.5^\circ} = 90$. Actually, for very high quality transmission it may be necessary to multiply from 500 to 1,000 times. It should be particularly noted that with this system of modulation the master-oscillator may be designed to possess the highest possible frequency constancy, since no attempt is made directly to vary its frequency. The frequency deviation must be symmetrical about the carrier frequency, and the simple balanced modulator described above will not fulfil this requirement. Referring to fig. 51 it is seen that if the side-band currents in T_a and T_b are equal, say i_a , the side-band voltages applied to the modulation amplifier will be proportional to $(\omega_r + \omega_a)i_a$ and $(\omega_r - \omega_a)i_a$, so that the vectors V_a , V_b will not be of equal amplitude. This inequality is easily removed by the insertion of a correcting network at a suitable point in the amplifying chain.

CHAPTER XII—PARAS. 86-89

86. Suppose that this system is applied to a transmitter which is normally driven by a master-oscillator at 2 Mc/s, with a frequency multiplication of 24, i.e. 3 doublers and a tripler, giving an operating frequency of 48 Mc/s. To adapt it to frequency multiplication, the new master-oscillator may be designed to operate at 61.25 kc/s, and the initial phase shift would be introduced at this frequency, a multiplication of 32 being necessary before coupling to the original transmitter; thus the overall multiplication would be $32 \times 24 = 768$. It is also of interest to note that changing the frequency of a frequency-modulated oscillation by the heterodyne method does not change the frequency deviation, e.g. if a frequency $f_r + f_a$ is caused to beat with a frequency f_o , the number of beats per second is $(f_r + f_a) \sim f_o$, say $(f_r - f_o) + f_a$. This enables an ingenious system to be used for a change of operational frequency.

87. If the above transmitter is required to change its operational frequency in the usual manner a considerable amount of readjustment would be required. If however the modulated master-oscillation is subjected to considerably greater frequency multiplication and then heterodyned by an oscillation of adjustable frequency, it is possible to dispense with a considerable amount of the readjustment. For example suppose the original oscillation (61.25 kc/s) to be multiplied 256 times, giving 15,680 kc/s. If this is heterodyned by an oscillator having a frequency of 13,680 kc/s, the output after rectification will be at 2,000 kc/s, and will still carry modulation with the original frequency deviation. The 2,000 kc/s oscillation may now be multiplied by 24 giving the final output at 48 Mc/s. If the final frequency is to be changed to 40.32 Mc/s, there is no necessity to interfere with the master oscillation or pre-heterodyne multiplier stages, for if the heterodyne oscillator is adjusted to 14,000 kc/s, the modulated beat oscillation will be 1,680 kc/s, and on adjusting the three doubling and one tripling stage to deal with this frequency the final output will be at 40.32 Mc/s. With this method of adjustment then, those stages which require the most careful manipulation can be set up once and for all, subsequent frequency adjustment being performed in the heterodyne oscillator and the four succeeding stages only.

Detection of frequency-modulated wave

88. The application of a frequency-modulated signal to any circuit network having a response dependent upon frequency will result in a wave which is modulated both in frequency and amplitude. Such a wave may be rectified in the same manner as a purely amplitude-modulated signal. The choice of a suitable circuit network presents a compromise between the conflicting claims of sensitivity and selectivity. If, for example, a frequency-modulated current passes through an inductance L of negligible resistance, the voltage v_L developed across its terminals is linearly amplitude-modulated. Thus if

$$\begin{aligned} i &= \mathcal{J} \sin \left(\omega_r t + \frac{\omega_r M}{\omega_a} \sin \omega_a t \right) \\ v_L &= -L \frac{di}{dt} \\ &= -\omega_r L \mathcal{J} (1 + M \cos \omega_a t) \cos \left(\omega_r t + \frac{\omega_r M}{\omega_a} \sin \omega_a t \right) \end{aligned}$$

i.e. the voltage wave is frequency-modulated and is also amplitude-modulated to a depth M . Since however in all practical cases M is small compared to unity this method is too insensitive to be of value as a means of translating pure frequency modulation into amplitude modulation.

89. Various methods of greater sensitivity have been proposed, all of which depend upon the shape of the resonance curve of a circuit possessing both inductance and capacitance. Suppose a frequency-modulated voltage of amplitude $\mathcal{E}$ to be applied to a series-resonant circuit as in fig. 52a. Let its mean frequency be $\frac{\omega_r}{2\pi} = f_r$ and the peak frequency deviation be $f_r M$. Then if the resonant frequency of the circuit is f_o , fig. 51b, the voltage across the inductance will have

the value $\mathcal{V}_0$, during periods of no modulation, while during the time sinusoidal frequency-modulating is occurring with a peak deviation $f_r M$ the voltage will vary between $\mathcal{V}_0 + \mathcal{V}_1$ and $\mathcal{V}_0 - \mathcal{V}_1$ as in the diagram, i.e. the grid-filament voltage will be amplitude-modulated. Such circuits are most effective when connected in a push-pull arrangement, in order that any slight amplitude modulation of the original signal shall be cancelled out, and the translation from frequency-modulation to amplitude-modulation may then be very nearly linear; it will not be entirely so since the slope of the resonance curve is not quite constant over the operating range. The distortion due to this is not serious unless the frequency swing is sufficient to carry the operating point over the peak of the resonance curve in which event distortion becomes so serious that speech is quite unintelligible. If, however, care is taken to limit the frequency deviation this method of detection appears to offer possibilities for aircraft communication. It should be noted that if the mean frequency is changed before detection from a high to a low value, as in a super-heterodyne receiver, the ratio of frequency deviation to mean frequency is increased, and a greater rectified current is obtained than if the detection is directly performed at the initial frequency.

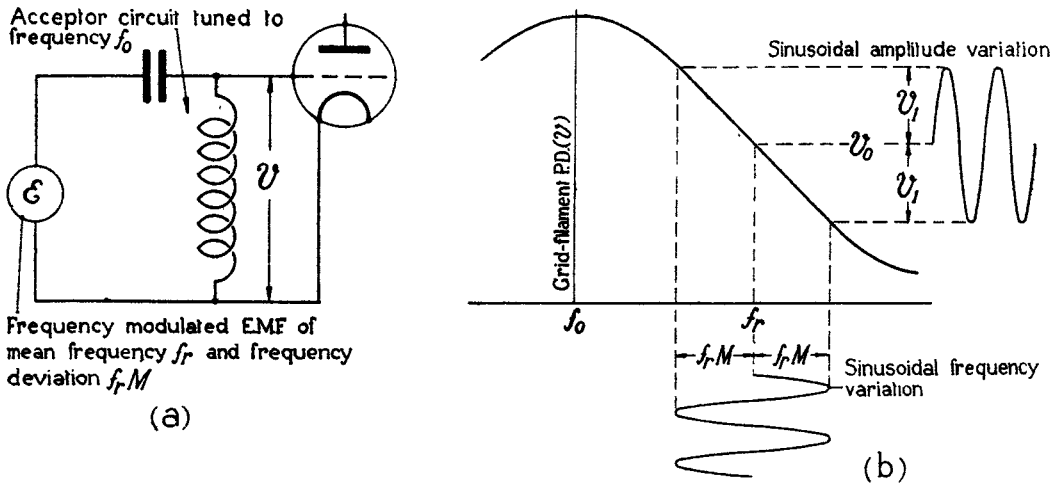


FIG. 52, CHAP. XII.—Translation of frequency modulation into amplitude modulation.

The type of receiver generally indicated is therefore a double super-heterodyne receiver in which the incoming very high frequency signal is first changed from the order of 100 Mc/s to say 5 Mc/s and afterwards to about 500 kc/s, thus providing considerable discrimination against second channel interference. Certain limiting stages must be introduced before the detector stage in order to eliminate any adventitious amplitude modulation which may be introduced either during propagation or in the tuned circuits of the receiver.

Signal-noise ratio and interference

90. It is claimed for frequency modulation (when carried out on very high frequencies with a large frequency deviation and a correspondingly wide frequency band) that the signal noise ratio is very much better than with amplitude modulation. In the latter case, the smaller the band of frequencies admitted by the receiver, the better is the signal-noise ratio, whereas in frequency modulation the greater the frequency deviation the wider is the band embraced by the side-frequencies and the greater is the depth of amplitude modulation after translation in the detector stage. Consequently, the signal-noise ratio increases with the frequency band embraced by the signal. For example, comparing an amplitude-modulated system with a frequency band of 20 kc/s with a frequency-modulated system having 100 kc/s frequency

CHAPTER XII—PARAS. 91-93

deviation for the same audio-frequency range, and assuming equal power to be drawn from the mains by both transmitters, the signal-noise voltage ratio of the frequency-modulated system is about 34 times that obtained by amplitude modulation. These are of course theoretical values, but it is claimed that improvements of this order are practically realizable. The actual improvement is however approximately inversely proportional to the initial noise level, i.e. the signal-noise ratio decreases as the noise level increases.

91. The possibilities of interference with frequency-modulated reception have not been completely investigated. The following results are probable, but are still to be verified experimentally.

(a) Amplitude-modulated signals should not interfere with frequency-modulated signals unless the two carriers are capable of combining to produce an audio-frequency beat note, i.e. the two carriers are nearly equal in frequency or separated by nearly twice the intermediate frequency.

(b) Second channel interference can occur between frequency-modulated signals exactly as between amplitude-modulated signals.

(c) It is claimed that two frequency-modulated transmissions can be separated, even if their side-bands overlap, provided that there is no audio-frequency beat between the carriers.

(d) No information is yet available with respect to cross-modulation.

Effect of propagation path

92. If an amplitude-modulated signal arrives at a receiver by two or more paths of different length, it undergoes what is called selective fading (Chapter XIV). In the same circumstances, a frequency-modulated signal is subjected to attenuation of some of the side-band frequencies to a greater extent than others, thus giving rise to distortion. Its use is therefore limited, either to short distances where only the direct ray is of importance, or to transmission at very high frequencies where there is no ray reflected from the ionosphere. The reflection at the surface of the earth must be taken into account, but at present there is little if any information available regarding the influence of these phenomena.

Summary

93. The present state of frequency modulation may therefore be summarized as follows :—

- (i) Under certain conditions of frequency and propagation path a frequency-modulated system provides quite a practical method of communication.
- (ii) Such frequencies are limited to the very high frequency band.
- (iii) Propagation must take place over a single path.
- (iv) The apparatus is rather more complicated than for an amplitude modulation system of comparable quality, but not unduly so.
- (v) The signal-noise ratio appears to be very much better than in the corresponding amplitude-modulated system.
- (vi) At the present stage of high frequency technique it should be possible to allocate as many frequency-modulated as amplitude-modulated systems within a given frequency band. This may not be true in the future, as the frequency stability of very high frequency amplitude-modulated systems may be considerably improved enabling allocation to be made much more closely than at present.

CHAPTER XIII.—RADIO-FREQUENCY MEASUREMENTS

The wavemeter

1. A wavemeter is an instrument which is used to adjust a transmitter or receiver to a desired frequency, or alternatively to measure the frequency to which the transmitter or receiver is adjusted. The term wavemeter arose from the former practice (now confined to very short waves and to discussions of aerial design) of referring to the wavelength of an oscillation rather than its frequency. Wavemeters used in the service may be divided into two classes :—

- (i) Absorption wavemeters.
- (ii) Heterodyne wavemeters.

The absorption wavemeter

2. This instrument is chiefly used to measure the frequency of the oscillation generated by a transmitter, and by a process of repeated readjustment to ensure that the desired frequency is radiated. It consists fundamentally of two portions, first a closed oscillatory circuit, the natural frequency of which may be varied over a certain range by variation of the value of inductance or capacitance, or both. For all possible settings of the adjustable component or components, the frequency of the circuit is accurately determined and recorded, either directly upon a scale attached to the variable condenser or inductance, or upon a chart which shows the frequency corresponding to any setting of the instrument. The circuit is therefore said to be calibrated. If only a comparatively narrow frequency band is to be covered, the oscillatory circuit may consist of a fixed value inductance and a variable condenser, while if a wide band is to be covered the inductance also may be adjustable by means of tappings. It is of interest to note that the residual capacitance (i.e. that of the circuit when the condenser is set to its minimum value), is generally of the order of one-tenth the maximum value of the condenser, so that with any inductance whatever, the frequency range is about 1 to $\sqrt{10}$ or say 1 to 3. It is generally desirable, in calibration, to ignore the first and last ten degrees of the scale, so that the useful range of an ordinary variable condenser is only 160° instead of 180° , and the frequency range covered with a given inductance is generally only of the order of 1 to 2.75, e.g. 500 — 1,400 kc/s. In certain designs, a continuously variable inductance is used in conjunction with a bank of fixed-value condensers.

Resonance indicators

3. The second essential portion of an absorption wavemeter is some device which will indicate the presence of an oscillatory current in the circuit, and its relative magnitude ; this may be called the resonance indicator. By this means the wavemeter may be adjusted to resonance with an oscillating circuit. In use the wavemeter is held in proximity to the latter, and its oscillation constant (i.e. L C value) is varied until the oscillatory current in this circuit is a maximum, as shown by the resonance indicator. The frequency of the oscillation is then identical with that to which the wavemeter is adjusted. The required degree of coupling between the oscillating circuit and the wavemeter depends upon the magnitude of the oscillatory current, and upon the sensitivity of the indicating device. To all intents and purposes, then, the absorption wavemeter is merely an extremely portable receiver of low sensitivity. A large number of indicating devices have been used, but it is only necessary to describe a few typical ones, namely :—

- (i) Incandescent lamp.
- (ii) Hot-wire ammeter or thermo-ammeter.
- (iii) Neon tube.
- (iv) Valve rectifier and micro-ammeter.

The incandescent lamp

4. This indicator was used in the No. 3 wavemeter formerly employed in aircraft. The lamp is usually of the miniature metal-filament type, rated at about .25 watts 2.5 volts, and its resistance is about 10 ohms when hot. It is connected either directly in series with the oscillatory

CHAPTER XIII.—PARAS. 5-7

circuit, or in an untuned circuit which is inductively coupled to the latter; as in fig. 1a and 1b respectively. Obviously, in order to raise the filament to its maximum safe incandescence, an E.M.F. of rather more than 2.5 volts R.M.S. must be induced in the oscillatory circuit. In use, however, the wavemeter should be withdrawn until the filament is barely glowing, when a much sharper resonance indication is obtained than when the lamp burns at full brilliancy. As a matter of interest, it may be noted that the resistance of the lamp is lower at low temperature than at high, and consequently the damping of the oscillatory circuit less, whereas if a carbon filament lamp is used the opposite is the case, hence the preference for a metal-filament lamp.

Current meter

5. A hot-wire ammeter or thermo-ammeter may be used in place of the lamp, provided it is sufficiently sensitive to give an appreciable deflection with a current of about 100 milliamperes. Such an instrument offers no advantage over a metal-filament lamp, and as it is expensive and delicate, this resonance indicator has no present service application.

Neon tube

6. The neon tube is without doubt the most generally useful of all indicating devices. It consists of a glass bulb in which are enclosed a pair of electrodes. The tube may be designed to fit an ordinary miniature bayonet joint or Edison screw lamp holder. The bulb is first evacuated

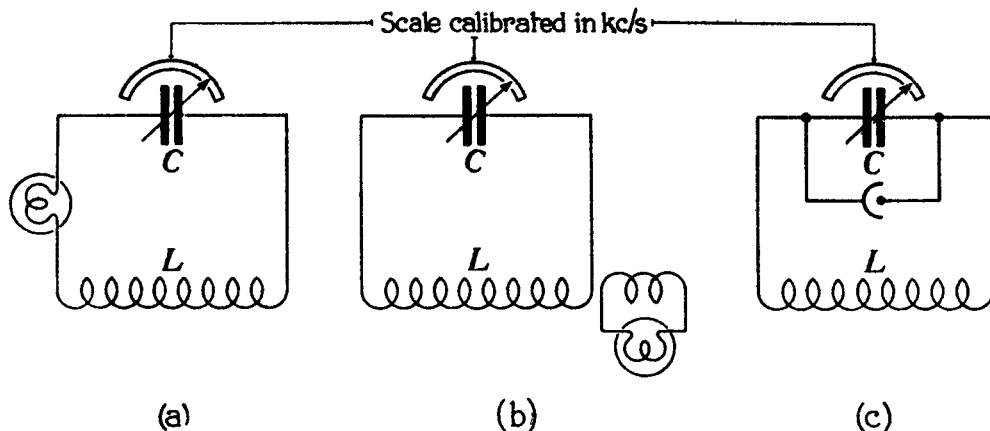


FIG. 1, CHAP. XIII.—Connections of incandescent lamp and neon tube for resonance indication.

to a high degree and a small quantity of neon gas is then introduced. When the P.D. between the electrodes exceeds about 180 volts, the gas becomes ionized and therefore partially conductive; a small convection current then flows between the electrodes, and the gas becomes incandescent, emitting a characteristic reddish-orange glow. Once the ionization has been initiated it will persist even if the P.D. between the electrodes is reduced to about 140 volts. As the tube is a potential-operated device it must be connected between two points which have a high P.D., i.e. in parallel with the capacitance of the wavemeter circuit (fig. 1c) and not in series as were the devices previously mentioned. The E.M.F. induced in the wavemeter must be of the order of 1 volt R.M.S., and with a circuit of moderately high magnification, say about 140, the condenser peak voltage will be rather more than 180 volts which is sufficient to "strike" the lamp.

7. All the above devices possess the disadvantage that the presence of the indicating device increases the damping of the oscillatory circuit, and consequently the sharpness of resonance. The discrimination of the wavemeter, that is, the smallest difference of frequency which can be detected, depends upon the sharpness of resonance of the circuit as well as the characteristics of the indicating device. As regards the latter, the neon tube gives better discrimination than any form of apparatus depending upon the heating effect of the current.

Valve rectifier

8. (i) The form of rectifying detector most suitable for use as a resonance indicator is a triode valve used as a lower anode-bend detector. The circuit of a typical instrument is given in fig. 2, in which L , C , constitute the oscillatory circuit, T is a triode, the anode circuit of which contains a microammeter A reading up to about 250 or 500 microamperes, and a H.T. battery of from 10 to 60 volts. The grid is maintained at a suitable negative potential with respect to the filament, so that the normal anode current is extremely small, e.g. one microampere. When an oscillatory E.M.F. is induced in the coil L , and a corresponding P.D. between grid and filament of the valve, the resulting changes of anode current are unsymmetrical; a rectified current then flows in the anode circuit, its magnitude being indicated by the meter. As the maximum grid-filament P.D. will be developed when the input circuit is tuned to the frequency of the inducing source, resonance is indicated by maximum reading of the microammeter. This resonance indicator possesses an important advantage over those previously described in that it

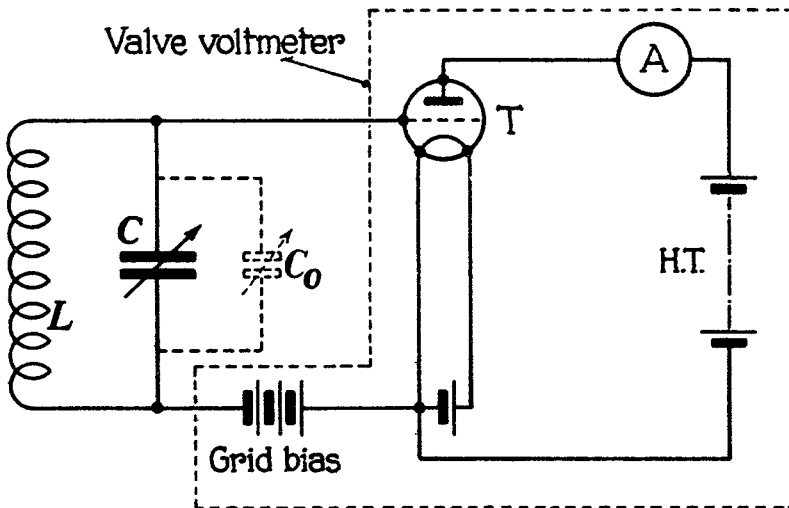


FIG. 2, CHAP. XIII.—Wavemeter with rectifier as resonance indicator.

imposes no appreciable damping upon the input circuit, provided that the anode circuit is of negligible impedance compared to the anode A.C. resistance of the valve. A small effective capacitance will be added in parallel with the input circuit, and this will not be the same for valves of slightly different characteristics. It is a simple matter to allow for this by including a small "trimming" condenser C_0 (fig. 2) in parallel with C so that slight adjustments may be made to compensate for an exchange of valves without alteration to the calibration. Where such a condenser is fitted, it must never be interfered with except by the person authorized to restore the calibration after such an exchange has been made. It is usual to supply two or more valves with the wavemeter, each bearing a serial number showing with which particular instrument they are to be used. These valves are so chosen that either may be used without affecting the calibration, and it is only necessary to adjust the "trimmer" when the initial supply of valves is exhausted. If care is taken to switch the filament off as soon as the wavemeter is finished with, the valve should have a very long life. All valves should be treated with care, but those supplied with wavemeters should receive particular consideration.

(ii) This type of resonance indicator is much more sensitive than any of the other types, a grid-filament P.D. of about 2 volts being sufficient to give full scale deflection, i.e. with an input circuit magnification of 100, the induced E.M.F. need only be .02 volt. In other words the sensitivity is about 100 times that of the neon tube while its discrimination is at least twice as good. When used for very high frequencies, a very open scale may be obtained by using a

CHAPTER XIII.—PARAS. 9-10

continuously variable inductance and a fixed condenser. A suitable type of inductance is that in which a radial arm is carried round each turn of the inductance in succession by a screw thread of coarse pitch (Chapter II).

Heterodyne wavemeter

9. The heterodyne wavemeter consists essentially of an oscillatory circuit which is maintained in oscillation by a thermionic valve, the triode being generally used. This wavemeter has a wide field of employment since it can be used to tune a receiver as well as a transmitter. When tuning a transmitter, the wavemeter is used as an autodyne receiver, and telephone receivers are introduced into the anode circuit of the valve, which acts as a detector in addition to maintaining the oscillation in its associated circuits. When used to adjust a receiver, the wavemeter acts as a transmitter of very low power, and having been adjusted to oscillate at the

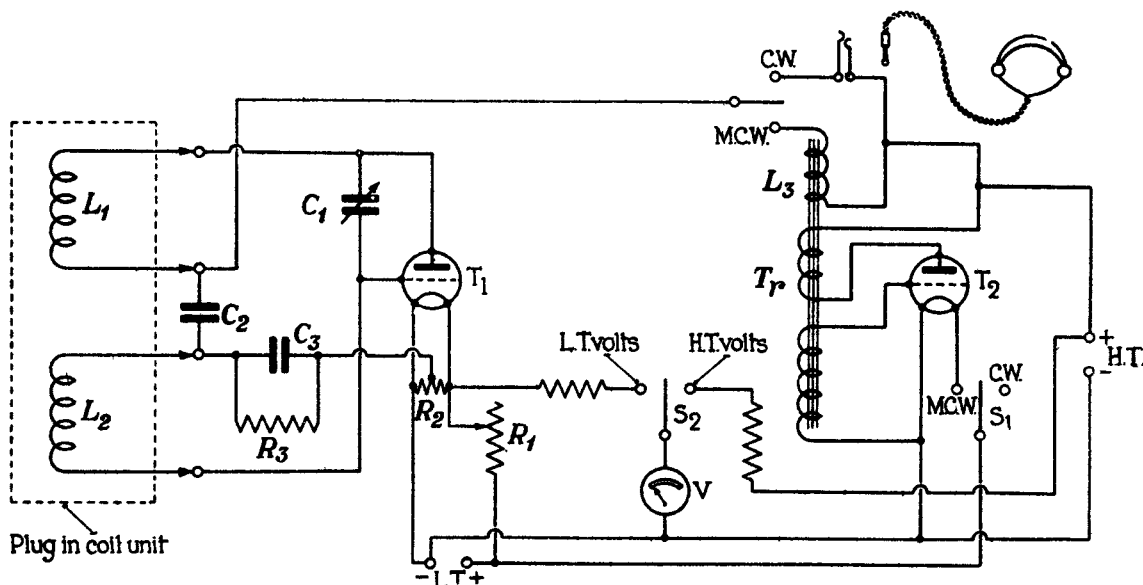


FIG. 3, CHAP. XIII.—Typical heterodyne wavemeter.

desired frequency, the receiver is easily adjusted to the latter. The circuit diagram of a typical heterodyne wavemeter is shown in fig. 3. Oscillations are maintained by the triode T_1 and the oscillatory circuit is of the Hartley type, consisting of any one of a series of plug-in coils, in order to cover a wide frequency range. The two inductances L_1 , L_2 which constitute this plug-in unit, are wound side by side upon a suitable former, and the whole winding is tuned by the variable condenser C_1 , while the condenser C_2 , which is of the non-inductive type and of large capacitance compared to C_1 , acts as a mains condenser. The grid is given a small initial positive bias by means of the potentiometer R_2 , and a grid leak R_3 and grid condenser C_3 are fitted in order to give a negative grid bias in the oscillatory condition. This combination ensures that the inception of oscillations occurs on the comparatively steep portion of the valve characteristic. As the oscillation increases in amplitude, however, the grid becomes negatively biased, and the grid-filament input conductance tends to assume a constant value. As stated in chapter IX this constancy is absolutely necessary if the frequency of the oscillation is to be unaffected by variation of L.T. and H.T. supply voltages.

10. The variable condenser usually has a maximum capacitance of about $0.0005 \mu F$, and its scale is graduated in degrees, a slow motion drive and vernier scale being often provided in order that the setting may be determined to 0.1 degree. The calibration of this scale, for each

particular plug-in coil, is recorded on a chart, the method of checking the calibration being detailed later. It is essential that the operating conditions under which calibration is performed should be reproduced as accurately as possible when the wavemeter is in use, and this entails that the H.T. voltage shall not differ by more than about 5 per cent. and that the filament emission should be maintained to an even higher accuracy. For this reason it is usual to fit a voltmeter by which both the filament voltage and H.T. voltage may be checked. In the instrument shown in the diagram the filament voltage is adjusted to a definite value, say 5 volts, by the rheostat R_1 , the voltmeter being connected in parallel with the filament, and the H.T. voltage is then checked while anode current is flowing by throwing over the switch S_2 . Suitable series resistances are fitted so that the voltmeter is direct reading for either of the two voltages.

Use of heterodyne wavemeter

11. (i) In tuning a transmitter, if exceptional accuracy is not required, it is sufficient to operate the transmitter while listening in the telephone receivers of the wavemeter. The condenser of the latter is then adjusted to the dead space and the frequency read from the calibration. In this process it is usually essential to remove the wavemeter to a considerable distance from the transmitter, otherwise the latter will pull the wavemeter oscillation into resonance and no heterodyne note will be heard. The width of the dead space is proportional to the E.M.F. induced by the transmitter, and by sufficiently reducing the coupling between transmitter and wavemeter, may be limited to about 200 cycles per second on either side of the resonant frequency. With the loosest possible coupling, a measurement made by the ordinary or zero-beat-note method may therefore be in error by about 200 cycles per second, no matter what the absolute frequency may be. The percentage error however is larger on low than on high frequencies.

(ii) When tuning an oscillating receiver, the same procedure may be followed, but it is generally more convenient to set the wavemeter in oscillation at the desired frequency and to tune the receiver to the wavemeter emission, just as if the latter were due to a distant high-power transmitter instead of an immediately adjacent, low-power one. In order that receivers designed for modulated-wave reception may be tuned, it is desirable to arrange for the wavemeter to emit modulated waves. In the instrument illustrated this is achieved as follows. The valve T_2 is arranged as an audio-frequency oscillator, its filament being heated when the filament switch S_1 is set to "M.C.W." Instead of the telephone receivers, the coupling coil L_3 , which is wound on the core of the audio frequency transmitter T_r , is now in series with the H.T. supply to the valve T_1 . An audio-frequency oscillatory flux is set up in the core of the transformer, and the anode voltage of the oscillator valve T_1 varies about its mean value. The amplitude of the radio-frequency oscillation varies correspondingly, and the action is to all intents and purposes the same as in choke control modulation (Chapter XII). Alternatively, the modulator valve may be dispensed with and the grid condenser and leak given such values that the "squegger" type of oscillation is maintained by the oscillator valve. The disadvantages of this device are first, the difficulty of choosing values for C_3 , R_3 , which will maintain this form of oscillation over a wide frequency range, and second, the calibration will not maintain a high degree of constancy under these conditions.

Tuning by double beat method

12. Provided that the initial accuracy of the wavemeter is sufficient, the percentage accuracy on low frequencies can be improved by utilizing what is called the double beat method of tuning. This requires the use of an additional oscillator, which need not be calibrated. The principle of the method is as follows. Suppose a transmitter to emit a frequency f_t and the wavemeter to emit a frequency f_w , then owing to the rectifier action in the wavemeter a beat note of frequency $(f_t \sim f_w)$ is perceived. If the additional oscillator emits a frequency f_o and the transmitter is inoperative, the beat note in the wavemeter telephones will be $(f_w \sim f_o)$ cycles per second. When the emissions of both transmitter and oscillator are simultaneously received, both these beat notes will be heard, and in addition a third beat note which is caused by the heterodyne effect of the two difference frequencies.

CHAPTER XIII.—PARAS. 13-14

13. (i) The procedure for adjusting the frequency of a transmitter to a predetermined figure by this method is as follows. Set up the wavemeter in accordance with its particular operating instructions, setting the frequency to the exact figure desired. A valve oscillator such as a syntonizer, or Receiver type R.64 with screening cover removed, is also adjusted for operation on the desired range. Listen in the telephones of this oscillator and adjust it to the dead space of the wavemeter emission. Press the key of the transmitter and adjust its frequency until a beat note is heard in the oscillator telephones, and bring this also into the dead space by adjustment of the transmitter. Do not vary the oscillator tuning during this operation. When this adjustment has been made, however, the second oscillator setting is varied slightly. As a rule, two distinct heterodyne beat notes will then be heard. By varying the transmitter frequency, these two notes are made to approach each other and a second beat will be established between them. This second beat becomes lower and lower in pitch and finally becomes zero, when the transmitter frequency is exactly that of the wavemeter. For example suppose the transmitter is to be tuned to 21,000 cycles per second. Then $f_w = 21,000$. Bring the oscillator frequency near to this, say $f_o = 22,000$, and adjust the transmitter as above to give a beat note with the oscillator, e.g. $f_x = 21,500$. We then have $f_w \sim f_o = 1,000$, $f_x \sim f_o = 500$. Now suppose f_x is shifted to 21,200 without changing f_w and f_o . Then $f_w \sim f_o = 1,000$, $f_x \sim f_o = 300$ and we get a second beat of 700. Shift f_x say to 21,100; $f_w \sim f_o = 1,000$, $f_x \sim f_o = 900$, and the second beat = 100. Shift f_x say to 21,020; $f_w \sim f_o = 1,000$, $f_x \sim f_o = 980$, the second beat = 20 and so on. Eventually we may adjust f_x to 21,001, $f_w \sim f_o = 1,000$, $f_x \sim f_o = 999$, and we get a second beat of one cycle per second. It is therefore possible after a little practice to adjust the transmitter almost exactly to the frequency of the wavemeter. There is a possibility that the second beat note may vanish owing to f_w and f_x being equally spaced above and below f_o , and to guard against this a check may be made by changing the frequency of the oscillator when the condition of zero double beat is obtained. If the transmitter and wavemeter are exactly in resonance, the single beat note will change but no double beat will re-appear.

(ii) To check the calibration of a wavemeter the double beat method may be used in combination with certain standard frequency transmissions, particulars of which are announced from time to time in Air Ministry Orders. The standard transmission takes the place of the wavemeter in the above explanation while the wavemeter is treated as the transmitter.

Tuning by harmonics

14. The range of a wavemeter may be extended to frequencies higher than those of its fundamental calibration by employing the harmonics of its oscillation. For this reason the circuit design is generally such that appreciable harmonics exist up to the 10th or even higher orders. Before attempting to use a wavemeter in this way for the first time, it is advisable to observe the relative intensity of the harmonics, e.g. by listening in the wavemeter telephones while rotating the tuning dial, in the neighbourhood of an oscillating receiver. As both oscillators produce harmonics, a large number of "chirps" are observed in this process, not only due to the fundamental of one oscillation and harmonics of the other but between say the second harmonic of one and the third, fourth, etc., of the other. A characteristic of these is the exceedingly minute portion of the scale in which the whole gamut of beat notes is perceived. The use of harmonics in tuning is best illustrated by a concrete instance. Suppose a transmitter to be in oscillation at a frequency which is known to be in the region of 18,000 kc/s, and the upper frequency limit of the wavemeter to be only 9,000 kc/s. Listening in the wavemeter telephones, note the successive readings at which "chirps" are heard. It is a good plan to sweep over the dial carefully and note approximately the location of each. Then repeat the process, approaching each point as slowly as possible, and taking full advantage of any slow-motion drive which may be fitted to the condenser, note the zero beat point of each harmonic as accurately as possible. Next tabulate the corresponding frequencies by reference to the calibration, as in the first two

columns of the table below. The successive frequencies at which the dead space of each "chirp" occurs should now be multiplied successively by a series of consecutive integers thus :—

- (1) 6,100 ($\times 2 = 12,200$)
- (2) 5,632 ($\times 3 = 16,896$)
- (3) 5,230 ($\times 4 = 20,920$)
- (4) 4,574 ($\times 5 = 22,870$)
- (5) 3,661 ($\times 6 = 21,966$)
- (6) 3,052 ($\times 7 = 21,364$)
- (7) 2,614 ($\times 8 = 20,912$)
- (8) 2,288 ($\times 9 = 20,592$)
- (9) 2,034 ($\times 10 = 20,340$)
- (10) 1,831 ($\times 11 = 20,141$)

The products shown in the right-hand column increase in value and then decline. This is an indication that beats caused by higher harmonics of the transmitter have been included in the series, frequencies (2) and (3) being suspect. The multiplication should be repeated, ignoring these.

- (1) 6,100 ($\times 2 = 12,200$)
 - (2) 4,574 ($\times 3 = 13,722$)
 - (3) 3,661 ($\times 4 = 14,644$)
 - (4) 3,052 ($\times 5 = 15,260$),
- etc.

This series is in an increasing progression throughout, showing that the wavemeter harmonics are of a higher order than those ascribed. If however the multiplication is repeated thus :—

- (1) 6,100 ($\times 3 = 18,300$)
- (2) 4,574 ($\times 4 = 18,296$)
- (3) 3,661 ($\times 5 = 18,305$)
- (4) 3,052 ($\times 6 = 18,312$)
- (5) 2,614 ($\times 7 = 18,298$)
- (6) 2,288 ($\times 8 = 18,304$)
- (7) 2,034 ($\times 9 = 18,306$)
- (8) 1,831 ($\times 10 = 18,310$),

the consecutive products now closely approximate to 18,300. The mean of the eight results gives the transmitter frequency as 18,304 kc/s, the average error being about 5 kc/s. The "chirp" at 5,632 kc/s is easily seen to be caused by the beating of the 4th harmonic of the transmitter (22,568 kc/s) with the 13th harmonic of the wavemeter, while that at 5,230 kc/s is due to the second harmonic of the transmitter (36,610 kc/s) and the 7th of the wavemeter. If the transmitter is to be adjusted as accurately as possible to 18,000 kc/s, the wavemeter should now be set to exactly 6,000 kc/s, and the transmitter carefully readjusted to give a zero beat at this setting. It is advisable to have another check run of at least five harmonics, and take the mean frequency as that of the transmitter. From the foregoing it is seen that some little care must be exercised in interpreting the results obtained by harmonic tuning. In practice however, the "chirps" caused by transmitter harmonics are generally weaker than those caused by its fundamental and they are easily rejected after a little experience.

Combination of heterodyne and absorption meters

15. A well-designed wavemeter of the neon tube or valve rectifier type is capable of much higher accuracy than can be achieved in a portable heterodyne wavemeter. In cases where a wavemeter of the former type is supplied for the purpose of transmitter tuning, the receiver can

CHAPTER XIII.—PARAS. 16-19

be tuned with the same degree of accuracy, provided that some form of heterodyne oscillator is available. The calibration of the latter is entirely ignored, and the procedure is as follows. Let us suppose it is desired to tune an oscillating receiver to a given frequency. First the transmitter is tuned to this frequency by means of the absorption wavemeter, using an artificial aerial where necessary. If the receiver is well screened, and fairly remote from the transmitter, it may be possible to tune the receiver directly to the transmitter, and nothing further is required. The transmitter and receiver, in situ, may however be interconnected in such a manner that the send-receive switch renders one or the other inoperative, and the heterodyne oscillator is then employed in the following manner. After tuning the transmitter to the desired frequency, the oscillator is syntonized with the transmitter by the dead space method, the telephone receivers being of course inserted in the heterodyne circuit. On switching over to "receive" the receiver may now be brought into resonance with the heterodyne oscillator in the usual manner, using the receiver telephones to obtain the dead space.

16. Where the receiver is of a non-oscillatory type, i.e. an R/T or I.C.W. receiver, either of the above methods may be used. If the receiver is to be tuned directly from the transmitter, the latter must be modulated by a source of sound of approximately constant amplitude, e.g. by fitting a small buzzer near the mouth-piece of the microphone, or even by holding a watch near it. If an additional heterodyne oscillator is employed, it must be tuned to the dead space as usual and subsequently, for the actual tuning of the receiver, must be caused to emit modulated waves.

17. An important application of the above expedient is to ensure that several aeroplanes which are to act in co-operation, are tuned to exactly the same frequency. This is achieved by tuning all transmitters with the same wavemeter, its frequency setting being rigidly clamped during transport from one aeroplane to another, and by tuning all the receivers to the emission of a single selected transmitter, generally that of the leader of the formation.

Double click method

18. An uncalibrated heterodyne oscillator may be brought into approximate resonance with an absorption wavemeter by what is called the double click method. The wavemeter is set to the desired frequency, and coupled to the oscillator. With the telephones in the latter circuit, if the heterodyne adjustment is varied, it will be found that a click is heard when resonance is nearly approached, and another click after passing the resonance point. These clicks are caused by sudden changes in anode current due to a fall in amplitude, or even, in the extreme case, to the cessation and re-establishment of oscillations. The reduction of amplitude is caused by the resistance load thrown into the oscillatory circuit of the oscillator by the presence of the absorbing circuit. When the coupling is very loose, this load is only small, the clicks are correspondingly faint and occur at closely adjacent points. With tighter coupling, so that the clicks are easily audible, they are separated by a greater distance on the scale. The resonant frequency is then taken as the mean of the two frequencies at which clicks occur.

Detection of clicks by current meter

19. (i) No great accuracy can be expected from the double click method under ordinary circumstances, first, owing to the difficulty of deciding the exact point at which each click occurs, and second, because to get clicks of good audibility, it is necessary to couple the two oscillatory circuits fairly closely. If it is possible to insert a sensitive moving coil instrument in either the anode or grid circuit of the heterodyne oscillator, the method becomes very sensitive and accurate. If a microammeter is available, it should be inserted in the grid circuit, as shown in fig. 4a, while if the only available instrument is a milliammeter, it should be inserted in series between the negative H.T. terminal of the oscillator and the negative terminal of the H.T. battery as in fig. 4b. The oscillator anode current should not exceed one-half of the maximum current for which the milliammeter is designed. The change of grid or anode current, at the point where the oscillation is wholly or partially suppressed, is shown by a violent flicker of the pointer. If the oscillator is fitted with a grid condenser and leak, the cessation of oscillations will be denoted

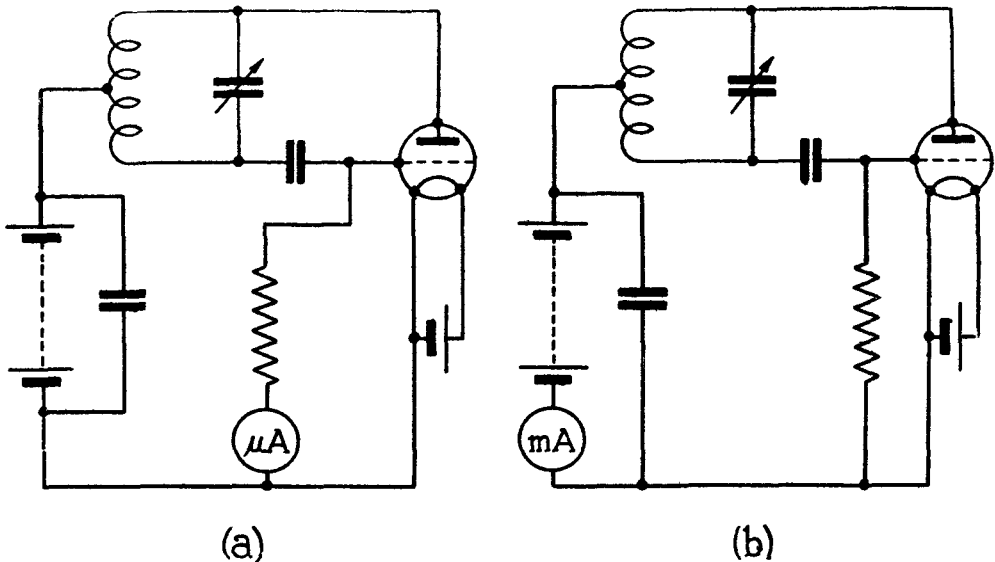


FIG. 4, CHAP. XIII.—Heterodyne oscillators with current meters as resonance indicators.

by a rise of anode current, and by a fall of anode current in the case of an oscillator without these components. Grid current usually falls on the cessation of oscillations in both types of oscillator.

(ii) The above information must not be regarded as an authority to interfere with the permanent wiring of a service wavemeter. It is intended to suggest a method of calibrating a heterodyne oscillator made up for the purpose of extempore measurements, such as are described in paragraphs 27 *et seq.*

The valve voltmeter

20. The resonance indicator described in paragraph 8, and enclosed in dotted line in fig. 2, is often referred to as a valve voltmeter, although strictly this term should only be applied if the microammeter scale is actually calibrated in volts. No calibrated valve voltmeter is at present standardized for service use, but where a suitable microammeter is available such an instrument is easily made up and calibrated. The voltmeter may then be used for the measurement of small radio-frequency or audio-frequency P.D.'s from 0.1 up to about 20 volts. Previous to the introduction of the valve voltmeter the measurement of such voltages was only possible in a well-equipped laboratory. The voltmeter can be so designed as to throw only a negligible load upon the circuit under investigation, and its indications are practically independent of frequency. It may therefore be calibrated at commercial frequency, using a potentiometer fed from the A.C. supply mains. Many different forms of instrument have been evolved for laboratory use, but only two will be dealt with here.

The anode-bend voltmeter

21. This is identical in principle with the resonance indicator previously described. The H.T. and grid bias voltages must be carefully chosen with respect to the valve and the particular anode current meter it is proposed to use. With most types of battery triode this form of voltmeter has a range of about 0 to 2 volts, and the grid bias should be about - 3 volts, in order to avoid grid current, which if allowed to flow will cause the meter to load the input circuit. The filament circuit should contain a rheostat with an off position, and care should be taken that the filament is never switched on unless there is a completely conductive path between grid

CHAPTER XIII.—PARAS. 22-23

and filament, in order that the correct bias is always applied to the grid. The H.T. voltage should be so chosen that a just perceptible anode current flows when the grid and filament are directly connected, and only about one quarter of the filament rheostat is in circuit. This deflection of the microammeter will then be the zero of the voltage calibration, the scale zero being disregarded. The meter may be calibrated by means of a potentiometer consisting of a length of 22 s.w.g. eureka wire, rather more than 36 inches being required. This is connected to an A.C. source, in series with a suitable adjustable resistance and a hot-wire ammeter, the resistance being so chosen that the current can be varied between say 0.5 ampere and 2.0 amperes. The potentiometer wire may be stretched on a bench or table, and by means of a Wheatstone bridge or bridge-megger, two points on the wire are located between which the resistance is exactly one ohm. Terminals are fitted at these points and the intervening wire divided into ten equal parts. The input terminals of the voltmeter are connected to the potentiometer via one of the terminals and a sharp edged prickler. It is obvious that with a current of one ampere in the

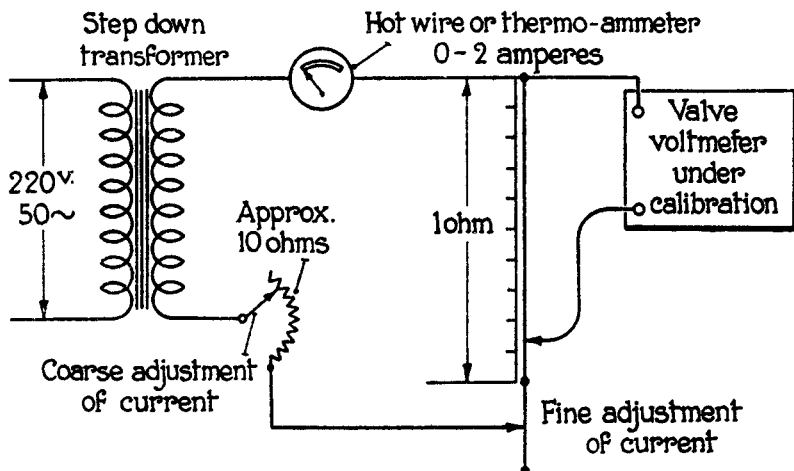


FIG. 5, CHAP. XIII.—Calibration of valve voltmeter.

wire, the P.D. between the terminals is one volt, and any desired fraction of this can be tapped off. The corresponding deflection of the microammeter is then noted, successive readings being plotted on a calibration curve. The scale may be extended from one to two volts by increasing the current in the potentiometer wire to two amperes. The method of calibration is shown diagrammatically in fig. 5.

22. When calibrated, it will be found that the scale is not uniformly divided, owing to the fact that the $I_a - V_g$ characteristic of the valve is curved. The linearity of the scale can be improved by the insertion of a non-inductive resistance of 30,000 to 60,000 ohms in the anode circuit. It is then necessary to connect a by-pass condenser of not less than $0.1 \mu F$ between anode and filament. It should have mica dielectric and be non-inductive. The whole instrument may be assembled on a board or in a box. Once the meter has been calibrated, the calibration may be relied upon to within about 5 per cent., provided that the operating voltages—filament, grid bias, and H.T.—are the same as when calibrating.

Peak voltmeter

23. If it is required to measure radio-frequency or audio-frequency voltages between say 5 and 50 volts, without attempting a very high degree of accuracy, the arrangement shown in fig. 6 may be adopted. The valve D is a diode, or triode with grid and anode in direct connection, and C is a condenser of high insulation resistance; for radio-frequency work its capacitance may be about $0.0005 \mu F$, and is not critical. Ignoring the presence of the resistance R and the

microammeter, it is seen that if an alternating voltage is applied to the input terminals, the condenser will charge until the P.D. between its plates is equal to the peak voltage at the terminals. With the resistance and microammeter connected as shown, this charge will leak away, and the average condenser P.D. is obtained from the product of the resistance R and the leakage current indicated by the microammeter. Provided the resistance is sufficiently large, the average condenser P.D. will only be a fraction of a volt below the terminal voltage. The instrument requires no calibration, but the value of the resistance should be so chosen that the voltage may be read directly from the microammeter scale by using a convenient conversion

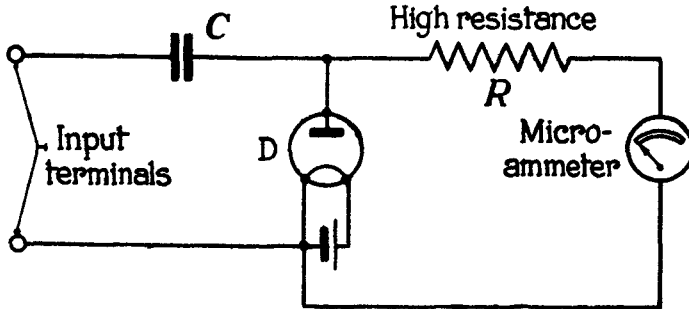


FIG. 6, CHAP. XIII.—Peak (diode) valve voltmeter.

factor. For example, if $R = 10^5$ ohms and the microammeter reads 0 — 200 microamperes, the average condenser P.D. is obtained by dividing the scale reading by 10, and the voltmeter has a range of 0 to 20 volts.

24. Although no great accuracy can be expected of the instrument when used in the above manner, it may be calibrated by the method described in paragraph 21. This calibration will of course give R.M.S. and not peak values. Owing to the incorporation of the reservoir condenser C , the calibration is not independent of frequency. It may be taken that if the supply voltage is E , and the condenser voltage V ,

$$V = \frac{\omega CR}{\sqrt{1 + \omega^2 C^2 R^2}} E.$$

Thus if $C = .001 \mu F$, $R = 10^5$ ohms, and the meter is used to measure the voltage E of the 50-cycle mains, it will read

$$V = \frac{2\pi \times 50 \times .001 \times 10^{-6} \times 10^5}{\sqrt{1 + (2\pi \times 50 \times .001 \times 10^{-6} \times 10^5)^2}} E \\ \doteq .03 E,$$

i.e. only one-thirtieth of the true voltage !

On the other hand, if a capacitance of $.2 \mu F$ is used

$$V = \frac{2\pi \times 50 \times .2 \times 10^{-6} \times 10^5}{\sqrt{1 + (2\pi \times 50 \times .2 \times 10^{-6} \times 10^5)^2}} E \\ = .98 E$$

i.e. an error of only 2 per cent.

25. If such a condenser were fitted permanently in circuit however, its comparatively large bulk would cause it to have appreciable capacitance to earth, and at high frequencies would constitute a path of low reactance in parallel with the grid and filament of the valve. This might introduce a serious error. The difficulty may be overcome by fitting a large condenser for

CHAPTER XIII.—PARA. 26

calibration at the supply frequency and substituting a small one for radio-frequency measurements. If the quantity ωCR remains the same the calibration will hold. For instance if $f = 50$, $C = .2 \mu F$, $R = 10^5$, $\omega C = 2\pi \times 10^{-5}$, $\omega CR = 2\pi$. At 10^5 cycles per second, the capacitance giving an equal admittance is

$$\begin{aligned} C_1 &= \frac{2\pi}{10^5 \omega} \\ &= \frac{2\pi}{10^5 \times 2\pi \times 10^5} \text{ farad} \\ &= .0001 \mu F. \end{aligned}$$

The chief disadvantage of this form of voltmeter is its low input impedance, which is rarely greater than $\frac{R}{3}$ and may be lower still. For this reason it is preferable to use a reservoir condenser of small capacitance, but the resistance R must then be correspondingly large in order to keep ωCR as large as possible.

26. An interesting application of the principle of the peak voltmeter is to measure the most positive potential reached by the grid of a radio-frequency power amplifier. In fig. 7 the

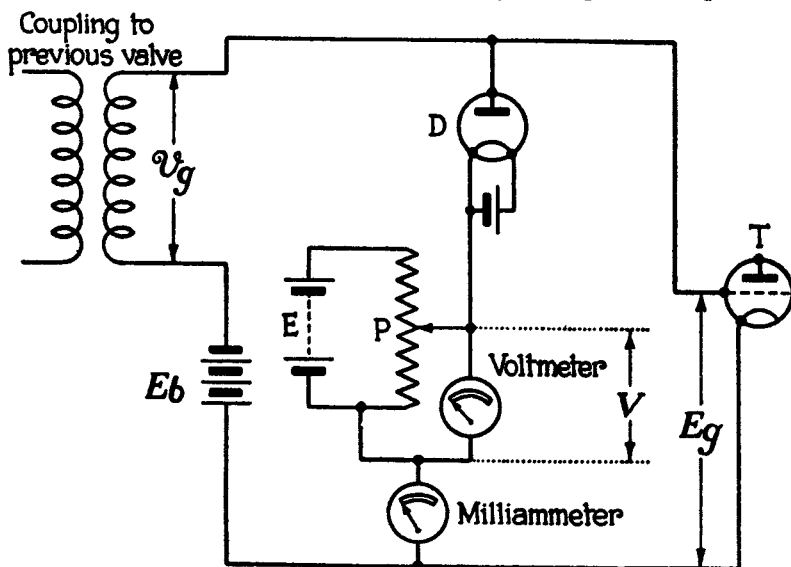


FIG. 7, CHAP. XIII.—Measurement of peak grid voltage of amplifier valve.

amplifier valve is denoted by T and it derives its excitation (of peak value $\mathcal{V}_g$) by means of an inductive coupling to the previous stage. The operating grid bias is supplied by means of a battery. The peak voltmeter consists of a diode D with its filament heating battery, a battery E, potentiometer P, moving-coil voltmeter and low-reading milliammeter. When the excitation reaches its positive peak, the grid of the triode is at a potential $E_g = \mathcal{V}_g - E_b$ volts above that of the filament (see Chapter XI). As the diode will pass current only if its anode is positive with respect to the filament, no current will flow through the milliammeter if the voltage V is greater than E_g . To measure the latter voltage therefore, the voltage V is reduced by means of the potentiometer until a very small current is detected, and the voltage is then again increased until the deflection of the milliammeter is reduced to zero. At this point, $V = E_g$. When used in conjunction with a potentiometer and battery in this manner, the diode is said to function as a "slide-back" voltmeter.

Measurements of inductance, capacitance and resistance

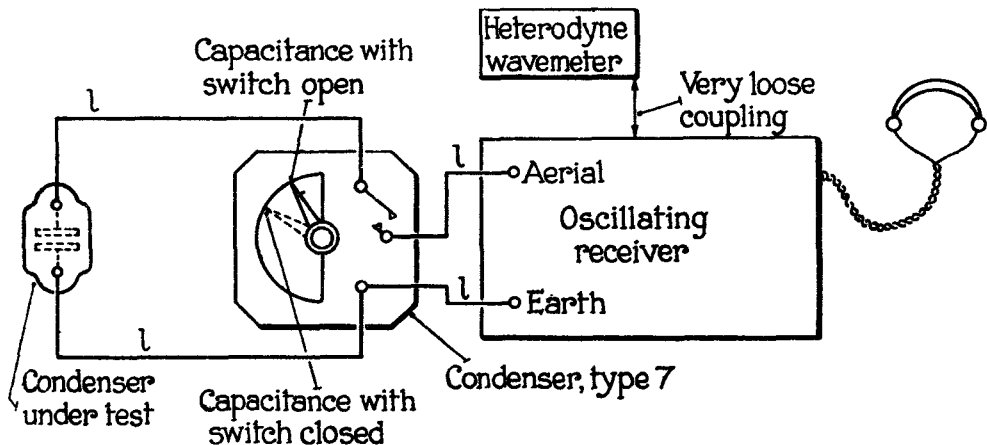
27. Occasions may arise, although rarely, when it becomes necessary to make measurements of these quantities; great accuracy is seldom necessary and it is not possible with the apparatus generally available. It is possible however to achieve a good deal with the aid of (i) a reliable heterodyne wavemeter. (ii) a good air-dielectric variable condenser with semi-circular moving vanes, e.g. the service type 7 condenser, and (iii) a valve voltmeter of the anode-bend type. The type 7 condenser is sometimes calibrated both in scale degrees, 0° to 180° , and in "jars."

The latter is a unit of capacitance equal to $\frac{1}{900}$ of a micro-farad, and is no longer used in the

Royal Air Force. The latter calibration is generally sufficiently accurate for any measurement which is required in the field. If the scale is marked only in degrees, it may be assumed that the minimum capacitance (scale reading zero) is $\cdot 00005 \mu F$, the maximum (scale reading 180°), $\cdot 00095 \mu F$, and that the capacitance varies uniformly between these limits.

Measurement of capacitance

28. To measure the capacitance of a small (fixed value) condenser, connect the terminals of the type 7 condenser as in fig. 8 to the aerial and earth terminals of a self-oscillatory receiver



NOTE: All connecting leads *l* to be as short as possible

FIG. 8, CHAP. XIII.—Measurement of capacitance of low value.

such as the type R.64, in place of the normal aerial and earth, and set up a heterodyne wavemeter or syntonizer to operate in the desired frequency range. The type 7 condenser should be adjusted to a capacitance rather larger than the estimated capacitance of the condenser under test, and the receiver set in self-oscillation. The wavemeter is then adjusted to the frequency of the receiver by tuning it to the dead space, using the wavemeter telephones for this purpose. The fixed condenser is then connected in parallel with the type 7 condenser, and the capacitance of the latter reduced until the receiver and wavemeter are again in resonance. The capacitance of the fixed condenser is equal to the change in capacitance of the type 7 condenser. In all measurements of this nature, it is essential to reduce the dead space to the lowest possible range by using a very loose coupling between the wavemeter and the tuned circuit which is under test. This test is of particular value in the case of an internal disconnection in a small fixed condenser, which is otherwise difficult to diagnose. An intermittent connection is also sometimes revealed by sudden variation in the beat note, and the presence of abnormal losses such as are caused by a high resistance connection to one or more electrodes may cause the receiver to stop

CHAPTER XIII.—PARAS. 29-30

oscillating when the condenser is connected in circuit. If an appreciable increase of reaction is necessary to maintain oscillation the electrical efficiency of the condenser should be regarded with suspicion.

Self-capacitance and inductance of a tuning coil or R.F. choke

29. To measure these, the coil is connected in parallel with the calibrated condenser to form a closed oscillatory circuit, and a valve voltmeter, which need not be calibrated in volts, is connected to the condenser terminals. A radio-frequency E.M.F. is induced in the circuit by loosely coupling to a heterodyne wavemeter, the arrangement being shown in fig. 9. It may be necessary to place the wavemeter several feet away from the oscillatory circuit, but in any case the coupling should be only just sufficient to give an appreciable scale reading in the voltmeter when the two circuits are in resonance. The standard condenser is set to say 170° and the

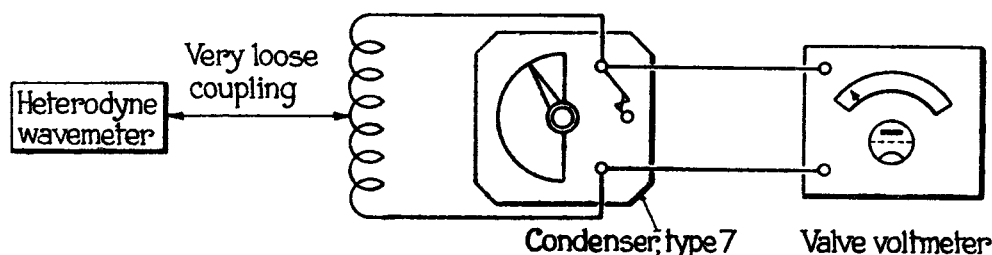


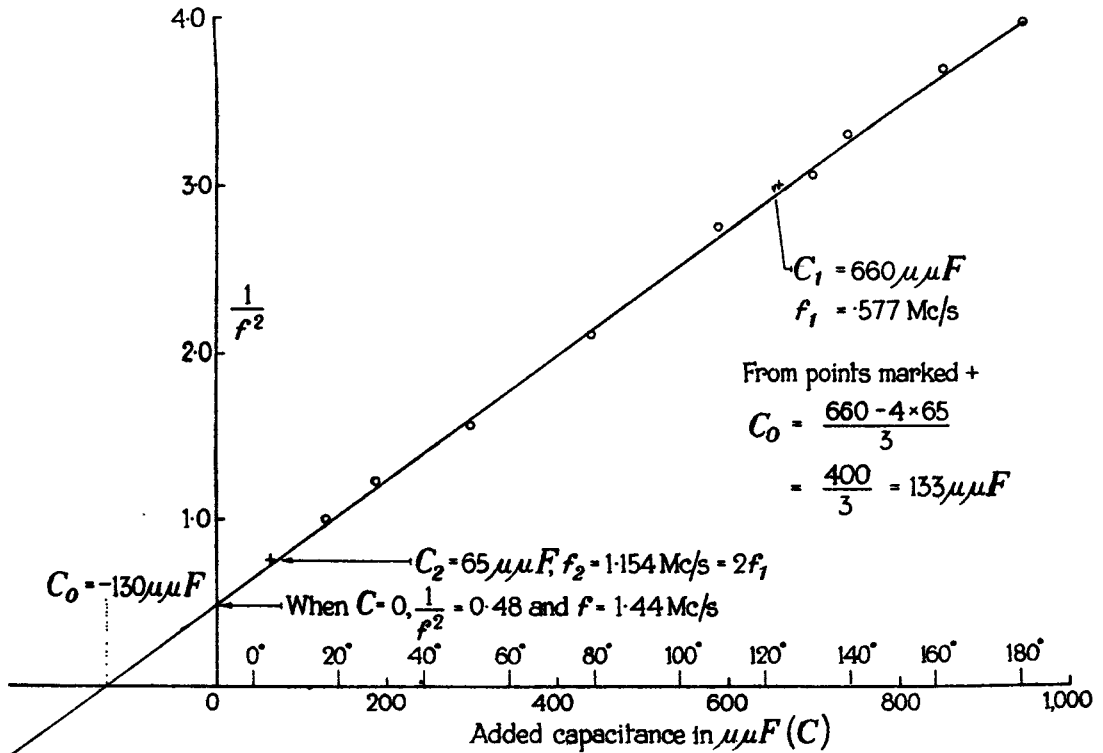
FIG. 9, CHAP. XIII.—Measurement of inductance and self-capacitance of coil.

wavemeter frequency is adjusted to resonance with the oscillatory circuit as shown by the voltmeter. The capacitance of the condenser and the wavemeter frequency are noted; the condenser is then re-set to say 160° , the wavemeter frequency adjusted to resonance, and again noted. The process is repeated for various values of capacitance and a table compiled as under:—

f (Mc/s).	$\frac{1}{f}$	$\frac{1}{f^2}$	Added capacitance C ($\mu\mu F$).
.5	2	4	940
.52	1.92	3.7	850
.55	1.83	3.3	740
.57	1.75	3.06	700
.60	1.66	2.75	590
.69	1.45	2.1	430
.8	1.25	1.56	300
.9	.81	1.235	190
1.0	1.0	1.0	130

30. The corresponding values of added capacitance C and $\frac{1}{f^2}$ are plotted as shown in fig. 10.

The points will be found to lie very nearly on a straight line, and the straight line lying most evenly between the points should be drawn. On extending this, it will be found not to pass through the origin; the intersection of this line with the $\frac{1}{f^2}$ axis gives the value of $\frac{1}{f^2}$ with no added capacitance, and from this the natural frequency of the coil alone is easily found. If the line is extended still further it cuts the capacitance axis on the negative side of the origin. This gives the value of capacitance which would exactly annul the self-capacitance of the coil, and

FIG. 10, CHAP. XIII.—Graphical derivation of L_0 and C_0

the self-capacitance C_0 is obviously equal in magnitude to this. In the diagram, the value of $\frac{1}{f^2}$, with no added capacitance, is .48, and the self-capacitance $130 \mu\mu F$. The inductance L_0 is calculated from these values, using the relation $f = \frac{10^6}{2\pi\sqrt{LC}}$ where L is in microhenries and C in microfarads. Let f_0 denote the natural frequency of the coil, then

$$f_0 \text{ (Mc/s)} = \frac{1}{2\pi\sqrt{L_0 C_0}}$$

$$f_0^2 = \frac{1}{2^2 \pi^2 L_0 C_0}$$

$$L_0 C_0 = \frac{1}{2^2 \pi^2} \times \frac{1}{f_0^2}$$

$$= \frac{1}{39.48} \times .48$$

$$L_0 = \frac{.48}{39.48} \times \frac{10^6}{130} \text{ (}\mu H\text{)}$$

$$= 93.5 \mu H.$$

CHAPTER XIII.—PARAS. 31-33

31. The disadvantage of this method is that the wavemeter has to be adjusted to resonance with the tuned circuit after each readjustment of added capacitance. The frequency must then be obtained from the calibration and its reciprocal is to be squared. Thus the arithmetical accuracy of the graph is generally poor. If the wavemeter has a pronounced second harmonic this disadvantage can be overcome as follows. First, add a fairly large capacitance C_1 to the tuned circuit, e.g. $670 \mu\mu F$, and find its resonant frequency; call this f_1 . Then, without disturbing the wavemeter setting, reduce the added capacitance while watching the valve voltmeter closely for a response to the second harmonic of the wavemeter frequency. If such a deflection is obtained, with a capacitance C_2 , the circuit is tuned to a frequency $f_2 = 2f_1$. Since the inductance L_0 is constant, it follows that

$$L_0 (C_1 + C_0) = 4 L_0 (C_2 + C_0)$$

$$\frac{C_1 + C_0}{4} = C_2 + C_0$$

$$\frac{C_1}{4} - C_2 = C_0 - \frac{1}{4} C_0$$

$$C_1 - 4C_2 = 3C_0$$

$$C_0 = \frac{C_1 - 4C_2}{3}$$

An example of this method is given in fig. 10. Here $C_2 = 65 \mu\mu F$, and the self-capacitance is found to be $133 \mu\mu F$. This value of C_0 , inserted in the formula $f_1 = \frac{1}{2\pi\sqrt{L_0 (C_1 + C_0)}}$ will give the value of the inductance.

32. It may here be mentioned that in the absence of facilities for measurement, the following empirical rule may be used for single-layer solenoids, viz., the self-capacitance of a coil, in micro-microfarads, is equal to $\cdot 64$ of the radius of the coil in centimetres. The natural frequency of the coil may be found by placing it in proximity to the heterodyne wavemeter, and listening in the telephones of the latter. As the wavemeter frequency is varied, a characteristic click should be heard when the natural frequency of the coil is passed through. On closer observation, it will be found that there are really two clicks, quite close together. They are caused by the reduction of amplitude of the wavemeter oscillation when the coil is in resonance, owing to the heavy damping then imposed upon the oscillator. The coil is in effect acting as an absorption wavemeter at this particular frequency. If the wavemeter is fitted with a sensitive meter in grid or anode circuit, resonance is indicated by a sudden flicker of its pointer (cf. paragraph 19).

Magnification of a closed oscillatory circuit

33. The magnification of a closed oscillatory circuit is easily determined by loosely coupling it to the heterodyne wavemeter. A calibrated valve voltmeter of the anode-bend type is connected across the condenser terminals of the oscillatory circuit, and the coupling adjusted so that the P.D. across the condenser is of some integral value near the upper limit of the voltmeter scale, e.g. 2 volts, when the wavemeter is exactly in resonance with the oscillatory circuit. Without altering the coupling, the wavemeter is then detuned until the condenser P.D. is reduced to $\cdot 707$ of its maximum value, in this case, $1\cdot 414$ volts. Let this be a frequency f_1 lower than the resonant frequency f_r . Sweep back through the resonant frequency, checking the peak value of voltage, and then detune to a frequency f_2 , higher than f_r , at which the condenser P.D. is again $\cdot 707$ of the maximum. The magnification x is then given approximately by the formula

$$x = \frac{\omega L}{R} = \frac{f_r}{f_2 - f_1}$$

Resistance of tuning coil or R.F. choke

34. If the coil is of such a shape that its inductance can be readily calculated, or can be found by the method previously explained, the resistance may be calculated from the magnification. If however it is required to measure the resistance independently, this can be performed as follows. It is first necessary to make up a set of sub-standard resistances, which must be of resistance wire of fine gauge, e.g. not less than 40 s.w.g. The resistance should be wound on a thin sheet of mica; a mica "card" about 2 in. by 4 in. may be used as a former, narrow slots about $\frac{1}{2}$ in. deep being cut in each edge. The wire is then wound in these slots until the desired resistance is obtained. This winding will have very low inductance owing to the closeness of the wires forming each complete "turn" of wire. Copper wires not thicker than 30 s.w.g. may be

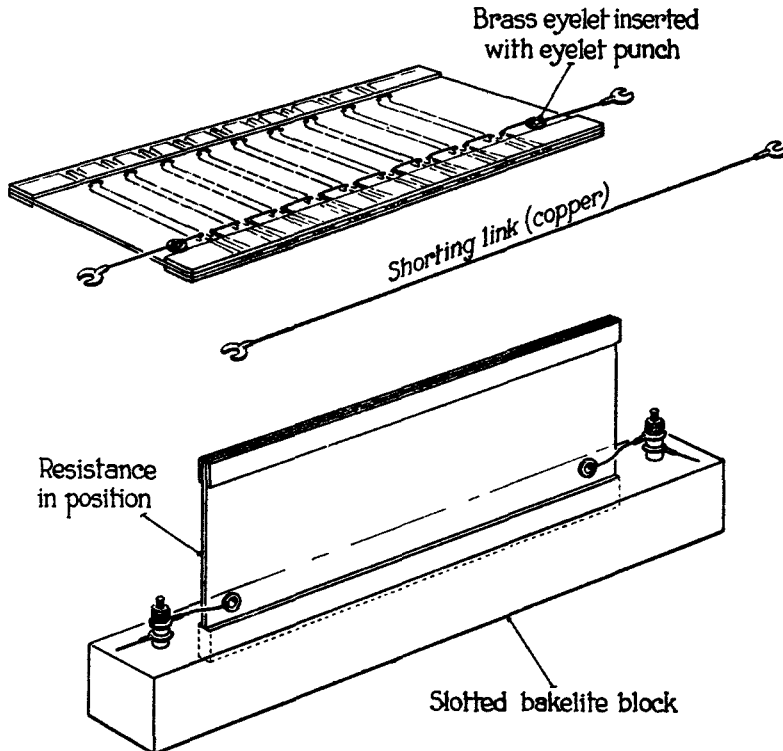


FIG. 11, CHAP. XIII.—Construction and mounting of sub-standard resistance.

used to connect the resistance in circuit, and a suitable holder is shown in fig. 11 together with a typical "card." The holder consists of a strip of bakelite slotted longitudinally to hold the card vertically, and the smallest obtainable terminals are used at each end to form binding posts for the interchange of resistances. The D.C. resistance of each should be measured by bridge-megger and it must be assumed that the R.F. resistance is the same as the D.C. resistance. This is the reason for stipulating that the finest available wire should be used. Eureka wire of 47 s.w.g. has a resistance of about 6 ohms per inch, while 40 s.w.g. has a resistance of about one ohm per inch. Where possible, of course, the resistance should consist only of a short straight wire. The approximate length for any given resistance may be found from Table I, Appendix A.

The resistance-variation method

35. The resistance-variation method of measurement is illustrated in fig. 12. The coil of which the resistance is to be measured, a condenser of high electrical efficiency (i.e. negligible

CHAPTER XIII.—PARA. 35

resistance), and the resistance holder described above are connected to form a closed oscillatory circuit. A calibrated valve voltmeter of the anode-bend type is connected across the condenser terminals, and a heterodyne wavemeter loosely coupled to the coil. The circuit is tuned to resonance with the wavemeter, the coupling being adjusted so that the voltmeter reading is near the maximum of its scale. During this operation the resistance holder is occupied by a short piece of copper wire, the resistance of which may be regarded as negligible. If E is the voltage induced in the circuit by the wavemeter, $\frac{\omega_r}{2\pi}$ the resonant frequency of the circuit, C the capacitance of the condenser, V_1 the voltage across the latter, and R_1 the resistance of the circuit

$$V_1 = \frac{1}{\omega_r C} \times \frac{E}{R_1}.$$

A known resistance R_2 is now connected to the resistance holder in place of the short copper link, and the circuit retuned to obtain maximum deflection in the voltmeter. It is important

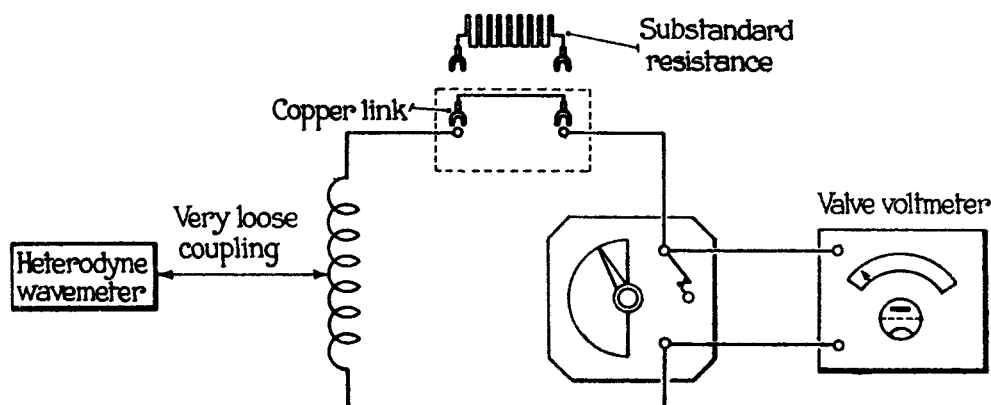


FIG. 12, CHAP. XIII.—Resistance-variation method of measuring R.F. resistance.

that the wavemeter, and the coupling between that instrument and the closed circuit, should not be disturbed in any way. On retuning, let the voltmeter reading be V_2 . Then

$$V_2 = \frac{1}{\omega_r C} \times \frac{E}{R_1 + R_2},$$

assuming that the introduction of the resistance has not changed the value of tuning capacitance.

From the above expressions, if $\frac{E}{\omega_r C}$ is denoted by K ,

$$V_1 = \frac{K}{R_1}$$

$$V_2 = \frac{K}{R_1 + R_2}$$

$$\begin{aligned} V_1 - V_2 &= K \left(\frac{1}{R_1} - \frac{1}{R_1 + R_2} \right) \\ &= K \left(\frac{R_2}{R_1 (R_1 + R_2)} \right). \end{aligned}$$

But

$$\frac{K}{R_1 + R_2} = V_2$$

$$\therefore V_1 - V_2 = V_2 \frac{R_2}{R_1}$$

$$R_1 = \frac{V_2}{V_1 - V_2} R_2.$$

It is advisable to check all measurements by adding at least two different known resistances.

36. The values of R_1 so obtained may differ from each other and the most convenient method of finding the probable value of R_1 is to plot $\frac{1}{V_2}$ against added resistance ; this procedure is analogous to that used in finding the self-capacitance of an inductive coil (paragraph 29). Suppose we obtain a set of readings as under :—

Added resistance (ohms)	V_2	$\frac{1}{V_2}$
0	2	·5
1	1·67	·6
2	1·43	·7
5	1·0	1·0
10	·67	1·5

On plotting $\frac{1}{V_2}$ against the added resistance, the points will be found to lie near a straight line, which should be drawn. On producing it to meet the "resistance" axis on the negative side of the origin, the intercept gives the value of negative resistance which will exactly annul the positive resistance of the coil, thus giving the value of the latter. In the above example, $R_1 = 5$ ohms.

37. The results obtained in this way may be very considerably in error owing to the effect of the self-capacitance of the coil. This error can be reduced by using a tuning capacitance very much larger than the self-capacitance. A correction may also be applied from the following considerations. If m is the ratio of the frequency f_r , at which the measurement is made, to the natural frequency f_o of the coil alone, i.e. if $m = \frac{f_r}{f_o}$, it can be shown that the measured value R_1 is related to the true resistance R by the equation

$$R_1 = \frac{R}{(1 - m^2)^2}.$$

Taking as an example the coil referred to in paragraph 29, the natural frequency of which was found to be 1·44 Mc/s, suppose the resistance is measured with 590 $\mu\mu F$ in parallel, so that $f_r = \cdot 6$ Mc/s.

$$m = \frac{f_r}{f_o} = \frac{\cdot 6}{1\cdot 44} = \cdot 416$$

$$m^2 = \cdot 173$$

$$(1 - m^2)^2 = \cdot 685$$

$$R_1 = \frac{R}{\cdot 685}$$

$$= 1\cdot 43 R$$

CHAPTER XIII.—PARAS. 38-39

If f_0 is known, as in the present example, the measured resistance may be corrected by using the above formulae. It is however preferable to make the measurement at a much lower frequency, e.g. $f_r = .25$ Mc/s. Then

$$\begin{aligned} m &= \frac{f_r}{f_0} = \frac{.25}{1.44} = .1735 \\ m^2 &= .03 \\ (1 - m^2)^2 &= .94 \\ R_1 &= \frac{R}{.94} \\ &= 1.065 R. \end{aligned}$$

The error is now only 6.5 per cent. If it is desired to keep the error due to self-capacitance alone to within 5 per cent., we must make $f_r \geq .158 f_0$. Then

$$\begin{aligned} m &= \frac{1}{6.32} \\ m^2 &= \frac{1}{40} = .025 \text{ (approx.)} \\ (1 - m^2)^2 &\geq (.975)^2 = .955 \\ R_1 &\geq \frac{R}{.955} \\ R_1 &\leq 1.05 R. \end{aligned}$$

The total error will be greater than this and may be of the order of 10 per cent.

38. A new difficulty now rises in that the measured resistance is that at the frequency f_r , which in the above case is 288 kc/s, whereas a coil of $92.5 \mu H$ is most likely to be used in a band in the region of 1 Mc/s. The trouble is prominent in this particular case because the self-capacitance is comparatively large. If it were only one-tenth of the given value, or $13 \mu F$, the natural frequency would be about 4.55 Mc/s and a measurement at .72 Mc/s, would require a correction for self-capacitance of about 5 per cent.

The reactance-variation method

39. This method obviates the necessity for a set of known resistances, but requires a calibrated variable condenser. The frequency at which the measurements are made must be known to a high degree of accuracy. The coil to be measured is connected as in the previous case (fig. 12) except that the resistance holder is omitted, and the wavemeter is again loosely coupled to the oscillatory circuit. The procedure in making a measurement is as follows. Set the wavemeter to a convenient frequency, $\frac{\omega}{2\pi}$, and tune the oscillatory circuit to exact resonance.

Note the capacitance C_r and condenser P.D., V_r , at this point. Now detune the circuit slightly, noting the new capacitance C_1 and condenser P.D., V_1 . Let the impedance of the de-tuned

circuit be $Z = \sqrt{R^2 + X^2}$. If E is the E.M.F. induced in the circuit by the wavemeter, and this is constant during the whole operation, we have

$$V_r = \frac{E}{\omega C_r R}$$

$$V_1 = \frac{E}{\omega C_1 Z}$$

$$\frac{V_r}{V_1} = \frac{C_1 Z}{C_r R}$$

$$= \frac{C_1}{C_r} \sqrt{1 + \frac{X^2}{R^2}}$$

$$\left(\frac{C_r V_r}{C_1 V_1} \right)^2 = 1 + \frac{X^2}{R^2}$$

$$\frac{X^2}{R^2} = \left(\frac{C_r V_r}{C_1 V_1} \right)^2 - 1$$

$$R^2 = \frac{X^2}{\left(\frac{C_r V_r}{C_1 V_1} \right)^2 - 1}$$

$$R = \frac{C_1 V_1}{\sqrt{(C_r V_r)^2 - (C_1 V_1)^2}} |X|.$$

Now $|X| = \omega L \sim \frac{1}{\omega C_1}$. When the circuit is tuned to the wavemeter frequency the reactance is zero, i.e.

$$\omega L = \frac{1}{\omega C_r}$$

Therefore

$$\begin{aligned} |X| &= \frac{1}{\omega C_r} \sim \frac{1}{\omega C_1} \\ &= \frac{1}{\omega C_r} \left(1 \sim \frac{C_r}{C_1} \right) \end{aligned}$$

and

$$\begin{aligned} R &= \frac{C_1 V_1}{\sqrt{C_r^2 V_r^2 - C_1^2 V_1^2}} \times \frac{1}{\omega C_r} \left(1 \sim \frac{C_r}{C_1} \right) \\ &= \frac{V_1}{\sqrt{\left\{ \frac{C_r}{C_1} \right\}^2 V_r^2 - V_1^2}} \times \frac{1}{\omega C_r} \left(1 \sim \frac{C_r}{C_1} \right). \end{aligned}$$

CHAPTER XIII.—PARAS. 40-41

40. This method is not very suitable for extempore measurements, because since R is generally only of the order of $\cdot 01 \times \frac{1}{\omega C}$, either $\frac{V_1}{\sqrt{\left(\frac{C_r}{C_1}\right)^2 V_r^2 - V_1^2}}$ or $1 \sim \frac{C_r}{C_1}$ must be

quite small. If the former is small, the voltage V_1 must be read on the cramped portion of the voltmeter scale, while if $\frac{V_1}{\sqrt{\left(\frac{C_r}{C_1}\right)^2 V_r^2 - V_1^2}}$ approaches unity, $1 \sim \frac{C_r}{C_1}$ will be of

the order of about $\cdot 01$, i.e. the change of capacitance only about one per cent. This is near the limit of accuracy of an ordinary commercial variable condenser. The difficulty can be overcome if in addition to the standard variable two standard fixed condensers are available. The capacitance (C_p) of one of these should be of the same order as that of the variable condenser

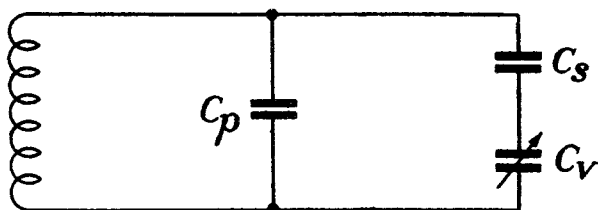


FIG. 13, CHAP. XIII.—Method of obtaining small range of frequency variation.

at 90° and that of the other (C_s) about one-tenth of this value. These are arranged as in fig. 13, from which it is seen that the total capacitance is $C_p + \frac{C_v C_s}{C_v + C_s}$. The total capacitance range is then from a little over C_p , when C_v is at its minimum, to a value approaching $C_p + C_s$ when C_v is at its maximum, i.e. the capacitance changes by rather less than the value C_s as the variable condenser is swung over its whole range. The complete condenser assembly is preferably calibrated in situ by means of a good capacitance bridge.

Capacitance bridge

41. (i) The principle of the Wheatstone bridge may be applied to the measurement of capacitance, provided that a standard condenser is available. The principle of the method is shown in fig. 14a in which R_1 and R_2 are two resistances of exactly equal properties, i.e. their resistances, inductances and self-capacitances are identical. The condenser under test is C_t , while C_s is a calibrated variable condenser. The correctness of balance is indicated by the absence of sound in the telephone receivers, which take the place of the galvanometer in the ordinary wheatstone bridge circuit, while instead of a steady E.M.F. from a battery, audio-frequency alternating E.M.F. is applied to opposite corners of the bridge. From ordinary circuit theory it is easily seen that when the capacitance of the calibrated condenser C_s is equal to that of the condenser C_t , the currents in the two parallel paths, i.e. R_1, C_s and R_2, C_t are equal, and the points X and Y at equal potential, hence no current will flow between X and Y and no sound will be heard in the telephones. When C_s is not equal to C_t however, this is not the case.

(ii) In a commercial form of this instrument, the alternating E.M.F. is produced by means of a small high-note buzzer of the Ericsson type, while the two ratio arms of the bridge are formed by the two halves of a differential condenser, and the instrument is given two ranges by the

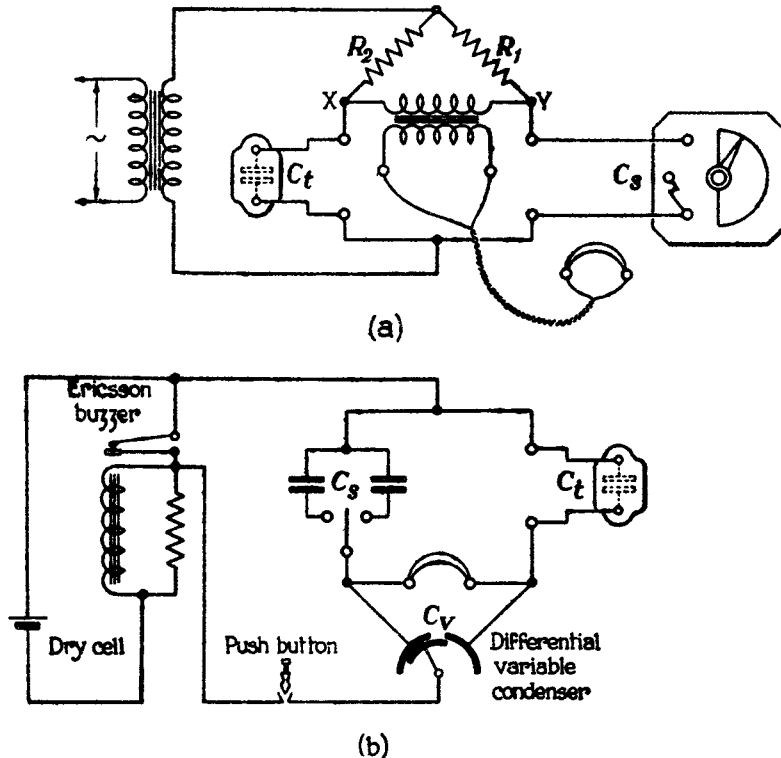


FIG. 14, CHAP. XIII.—Capacitance bridge.

adoption of two alternative fixed condensers in the arm opposite to that containing the condenser under test (fig. 14b). By this means a very wide range of capacitance can be covered in a single sweep of the differential condenser. The latter may be provided with two direct-reading scales of capacitance, and the value of C_t is read from the one corresponding with the particular standard condenser which is switched in for the purposes of the measurement.

Measurement of depth of modulation

42. Instruments used for determining, approximately, the depth of modulation of an R/T transmitter are known as modulation indicators. The principle of the method is shown in fig. 15. Here $L_1 C_1$ is a circuit tuned to the radiated frequency and loosely coupled to the aerial circuit of the transmitter, so that the condenser P.D. is of the order of 10 volts or more; the greater the amplitude the more accurate the result will be. The alternating voltage across the condenser is rectified by the diode D_1 , a reservoir condenser and load resistance C_2 , R_2 , being fitted as usual; the values of C_2 and R_2 are chosen in such a way that C_2 can be regarded as a short circuit for the radio-frequency current and of infinitely high reactance for an audio-frequency current while R_2 is so chosen that the time constant $C_2 R_2$ is of the order of 5. Suppose the transmitter to be sinusoidally modulated, so that the envelope of the condenser P.D. has the form of fig. 15b and f is the modulation frequency. The maximum value P and minimum value Q of the envelope can be measured by means of the diode D_2 , the microammeter M and the battery E together with the potentiometer R_3 .

43. To measure the maximum voltage P it is only necessary to put the switch arm in the position marked P (fig. 15a) and vary the adjustment of the potentiometer R_3 until the current

CHAPTER XIII.—PARA. 43

through the microammeter is zero. The reading of the direct current voltmeter V then gives the voltage P (fig. 15b). The change-over switch is then moved to the position Q and the measurement is done in exactly the same way, the reading of the voltmeter V now giving the minimum voltage Q (fig. 15b). When the microphone of the transmitter is quiescent the two readings of the voltmeter should be the same for both positions P and Q of the switch. Telephones are fitted in series with the microammeter and it will be observed that these telephones are not

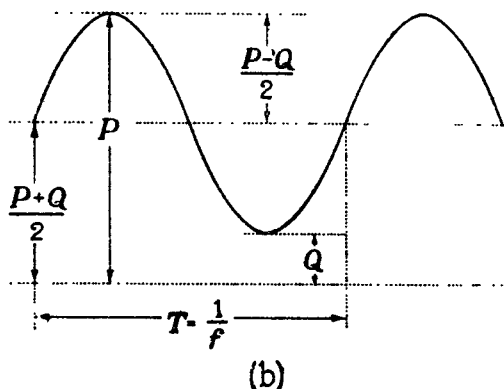
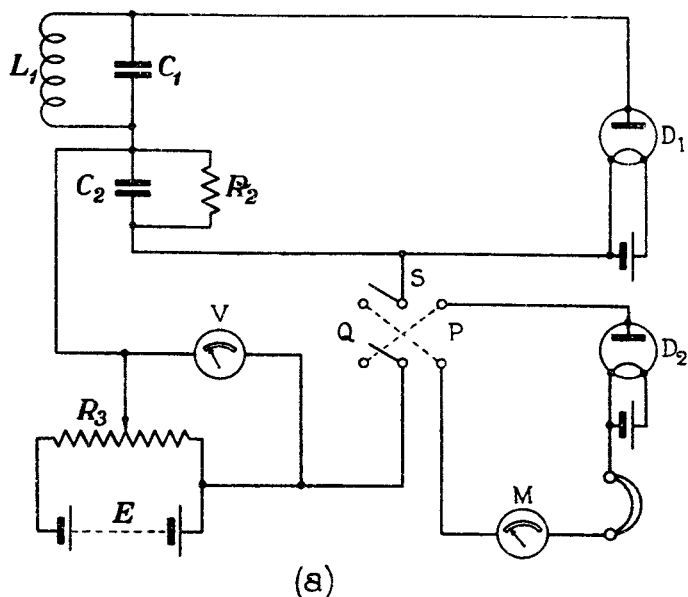


FIG. 15, CHAP. XIII.—Principle of modulation indicator.

absolutely silent at the moment when, by means of the potentiometer R_3 , the microammeter reading is brought to zero, because the diode passes some audio-frequency current owing to its capacitance. As soon as the micro-ammeter begins to show a current, a peculiar crackling noise will be heard in the telephones owing to peaks of the modulation being passed by the diode D_2 . The moment when the microammeter begins to pass the slightest current can be recognised by the change in the characteristic sound in the telephone, so that in practice the microammeter can be dispensed with.

44. Having obtained the maximum value P and minimum value Q from the voltmeter V , the depth of modulation K may be determined from the formula

$$K = \frac{P - Q}{P + Q}$$

Many modifications to this method have been made with the object of reducing the amount of manipulation required, but the advantage of the scheme outlined above is that it does not depend upon a prior calibration. If desired, the coil L_1 may be coupled to a radio-frequency amplifier and the measurement performed upon the transmission of a distant station, provided the gain is sufficient to give the required P.D., i.e. not less than 10 volts.

CHAPTER XIV. PROPAGATION OF ELECTRO-MAGNETIC WAVES

INTRODUCTORY

Frequency range of electro-magnetic waves

1. This chapter, dealing with the propagation of radio waves, may be divided into two portions, the first containing an outline of elementary theory leading to a brief and approximate treatment of the influence of ions and electrons upon the passage of an electro-magnetic wave, while the second gives a factual summary of the propagation of waves of different frequencies. In Chapter VII it was explained that when electrons are accelerated an electro-magnetic wave is produced. This wave consists of a sinusoidally varying electro-magnetic flux travelling with

a velocity of approximately $\frac{3 \times 10^{10}}{\sqrt{\mu \kappa}}$ centimetres per second, its velocity thus depending upon the magnetic permeability and the dielectric constant of the medium. Reference was also made to the reception of light and heat energy from the sun, and it was stated that this energy is conveyed in the form of electro-magnetic waves. Heat and light radiation is therefore of the same nature as that used for wireless communication, differing only in frequency; the whole spectrum of frequencies occupied by known forms of electro-magnetic radiation ranges from 10^{31} kc/s downwards. The known range of frequencies is shown in fig. 1, the frequencies being divided into bands according to the mechanism of production; certain of these bands overlap because the particular frequencies can be generated by two different methods.

2. The highest frequencies so far detected are the so-called cosmic rays, which appear to be produced far out in space, and reach the earth from all directions. They are capable of penetrating very deeply into metallic bodies, although the latter are practically perfect reflectors for waves of lower frequencies. The next broad band embraces the gamma-rays emitted by radio-active substances such as uranium, and X-rays, which are produced by the bombardment of a metal plate by electrons moving with extremely high velocity. The X-ray tube is in fact nothing more than a special form of high-vacuum diode. The lower-frequency X-ray band overlaps the band embraced by ultra-violet light, which is emitted by bodies at extremely high temperatures. Both X-rays and ultra-violet rays affect a photographic plate, but do not give the sensation of vision. They are of medical value, but are capable of inflicting severe injury upon the delicate structure of the human eye.

3. The next band—the spectrum of visible light—is that with which we are most familiar. When the whole visible spectrum is present, as in the radiation from the sun, the various frequencies combine in their action upon the eye and nervous system to give the sensation which we identify as white light. The lowest frequencies in this band give the sensation of red light, and the adjacent band, which embraces the radiation emitted by hot bodies at a temperature just below incandescence, is known as infra-red radiation. These rays are of course also emitted by incandescent bodies, e.g., of the total radiation from an electric lamp filament, about 99 per cent. is infra-red and only 1 per cent. visible light. It may be noted that the whole of the frequency range between about 10^{16} and 10^9 kc/s can be produced in an indirect manner by electrical means. Radiation of frequency below about 10^9 kc/s can be directly produced only by electrical oscillators, and can be detected only by electrical methods.

4. (i) Before dealing with the range of frequencies which for want of a better term we may call the radio-communication band, it may assist the correlation of ideas if it is pointed out that in general, we suppose all the forms of indirectly produced radiation to be initiated by large numbers of minute oscillators somewhat resembling miniature hertzian doublets (Chap. VII). Little is known about the actual mechanism of production of this kind of radiation, but a very rough idea of the production of, say, an oscillation in the ultra-violet region may be obtained

CHAPTER XIV.—PARAS. 5-6

as follows. Consider a single hydrogen atom, the radius of which is about 10^{-8} centimetres. Suppose that owing to the sudden impact of a free electron, the electron belonging to the hydrogen atom is disturbed in its orbit and executes a to-and-fro vibration, superimposed upon its normal motion. We then have two equal electric charges of opposite sign, alternately appearing at either end of an element of space about 10^{-8} centimetres in length. This may be compared with the oscillation of the charge on the plates of the hertzian doublet, and the disturbed atom may, on this hypothesis, be regarded as an elementary half-wave aerial. The wavelength will then be 2×10^{-8} centimetres, and the frequency $\frac{3 \times 10^{10}}{2 \times 10^{-8}} = 1.5 \times 10^{18}$ cycles per second or 1.5×10^{12} Mc/s; thus the disturbance of an atom in the manner indicated would give rise to a momentary emission of ultra-violet light. It will of course be appreciated that the actual mechanism of emission is far more complicated than is here implied.

(ii) The frequencies actually used in radio-communication are also shown in fig. 1. As already stated, they cannot be detected by the unaided human senses. At the present time, the practical upper and lower limits may be taken as 3×10^7 kc/s and 10 kc/s respectively. This band is sub-divided as follows:—

Below 30 kc/s	..	..	..	..	..	..	Very low frequencies (VL/F)
30–300 kc/s	..	..	..	..	..	..	Low frequencies (L/F)
300–3,000 kc/s	..	..	..	..	..	..	Medium frequencies (M/F)
3–30 Mc/s	..	..	..	..	..	..	High frequencies (H/F)
30–300 Mc/s	..	..	..	..	..	..	Very high frequencies (VH/F)
300–3,000 Mc/s	..	..	..	..	..	..	Decimetre waves (dc/W)
3,000–30,000 Mc/s	..	..	..	..	..	..	Centimetre waves (cm/W)

The present trend of development is towards the gradual abandonment of frequencies below about 100 kc/s, and exploration of the utility of the 300 to 30,000 Mc/s band. In the diagram, the radio-communication band has been extended to 10^9 kc/s, the latter being the highest frequency produced in the laboratory by direct electrical methods. There is, however, no present prospect of the practical production and employment of such high frequencies.

ELEMENTARY PROPAGATION THEORY

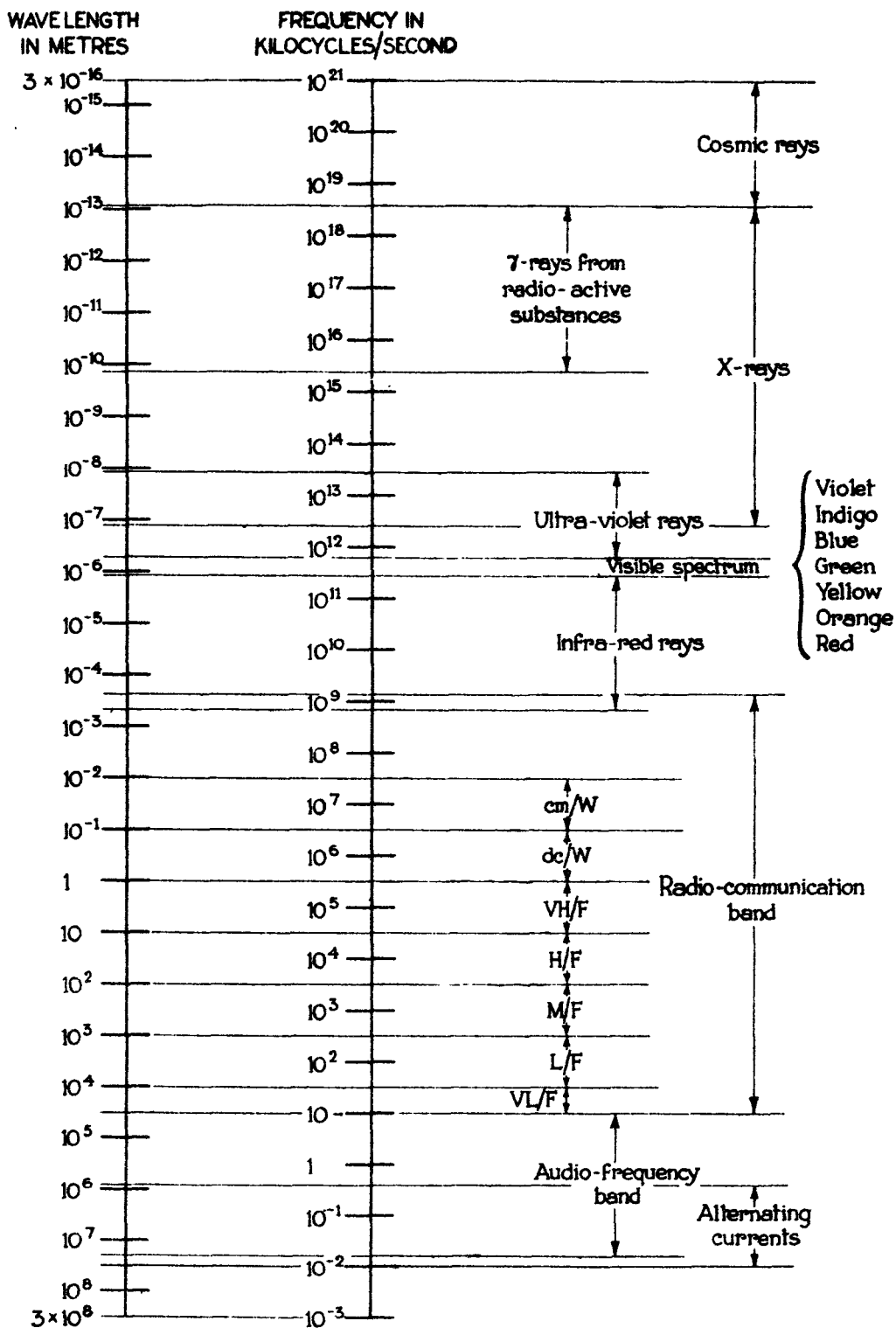
5. Since light, heat and radio waves are all of the same nature, it is to be expected that the same general laws govern their propagation through space, but the presence of material molecules, or even of free electrons, in the region through which propagation takes place, has a considerable effect upon the wave, its exact nature depending to a great extent upon the frequency. In order to explain the general laws, it is convenient first to consider light waves, because they are easily produced and are perceptible to the human senses without the aid of extraneous apparatus, so that many of the laws are easily verified by simple experiments. A ray of light is defined as the path along which a wave travels, and a collection of light rays is called a beam. It is a matter of every-day observation that light travels in straight paths; for instance, if a small hole is made in the blind of a darkened room, the light entering by this hole will illuminate the dust particles in the atmosphere and the light reflected from these enters the eye. It is sometimes then said that a beam of light is seen, but actually the only light seen is that reflected from the particles, and these are observed to lie in straight lines. A natural consequence of the rectilinear propagation of light is the formation of shadows, because the presence of an obstacle in the path of the beam may prevent the illumination of particles on the side of the obstacle remote from the source of light. Bodies are said to be transparent, translucent or opaque according to the degree to which they allow the light to pass through them.

Definitions

6. When light comes into collision with a material body three effects may be observed.

(i) A portion of the light is reflected at (or near) the surface of the body.

(ii) A second portion may travel through the body, the rays being usually bent on passing through the boundary surface. The rays are then said to be refracted.



ELECTRO-MAGNETIC WAVES IN FREE SPACE

(iii) A portion of the refracted light is absorbed during its passage through the body. In speaking of these phenomena, it is necessary to define

(a) The normal to the surface of the body. This is a line drawn through and perpendicular to the surface, at the point of incidence of a light ray.

(b) The incident ray, which is the path in which light travels before reaching the surface.

(c) The reflected ray, which is the path taken by the light after it has collided with the surface.

(d) The angle of incidence, or angle which the incident ray makes with the normal.

(e) The angle of reflection, or angle which the reflected ray makes with the normal.

(f) The angle of refraction, or angle which the refracted ray makes with the normal.

Reflection

7. The laws of reflection are as follows :—

(i) The angle of reflection is equal to the angle of incidence.

(ii) The incident ray, the reflected ray and the normal through the point of incidence all lie in the same plane.

These laws are illustrated in fig. 2.

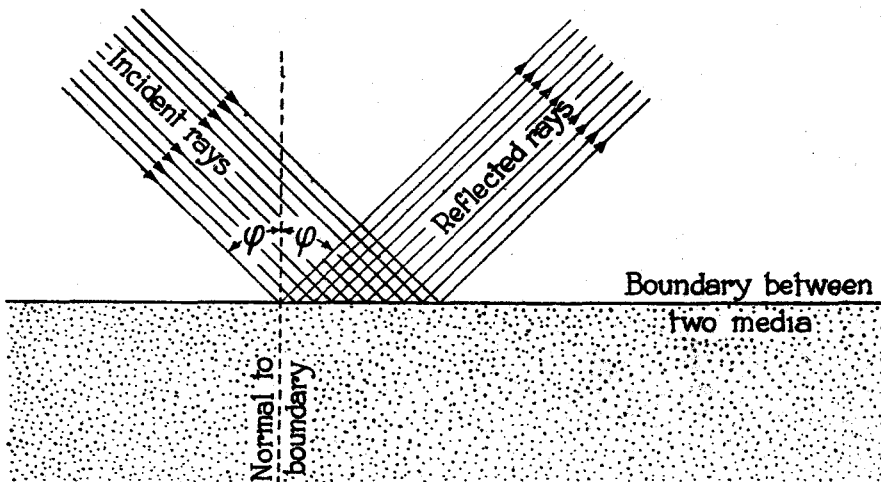


FIG. 2, CHAP. XIV.—Reflection of electro-magnetic wave.

Diffraction

8. In the early history of the wave theory of light, it was strongly opposed by Newton, because it appears that light travels in straight lines, whereas sound, which had previously been proved to be a wave motion in a material medium, was found to bend round obstacles. The scientist Fresnel appears to have been the first to point out that a light ray does in fact bend round an obstacle just as a sound wave does, the difference being merely one of degree. The bending of a wave round the edges of an obstacle is referred to as diffraction. It may be observed by allowing light to pass through a very small aperture and to fall on a screen. If an obstacle is interposed so as to cast a shadow upon the screen, the edges of the shadow are never perfectly defined. Close observation reveals that several alternately light and dark bands appear outside the shadow, and these get nearer to each other as they recede from the edge of the shadow, until they merge into the area of uniform illumination. Another example of diffraction may be

CHAPTER XIV.—PARAS. 9-10

observed by viewing a distant source of light, such as a street lamp, through a mist or light fog, when the lamp itself appears to be surrounded by rings of coloured light. This is due to the bending of the light rays round the particles of water vapour, and it is found that the lower frequency (red) rays are bent to a greater degree than those of higher frequency. It is in fact capable of demonstration that the lower the frequency, the greater the degree of bending. Diffraction phenomena are therefore of considerable importance at radio-communication frequencies.

Refraction

9. The term refraction refers to the bending which takes place when a ray passes from one transparent medium to another. The effect is easily shown by the partial immersion of a stick in water, when the stick appears to be bent at the point where it meets the surface. The relation between the angle of incidence and the angle of refraction is known as Snell's Law, and is as follows :—

When a ray of light is incident upon the surface of separation of two media, it is bent in such a manner that the ratio $\frac{\text{sine of the angle of incidence}}{\text{sine of the angle of refraction}}$ is a constant for all angles of incidence. This constant is called the refractive index of the two media, and is denoted by the symbol n . In optics the symbol μ is usually used for the refractive index, but has the disadvantage of confusion with the magnetic permeability of the medium.

Physical meaning of n

10. The refractive index has a real physical meaning apart from the geometrical one just described, for it depends upon the velocity of light in the two media. It is now convenient to refer to the wave-front of a wave, which is defined as an imaginary surface perpendicular to the direction of propagation of the wave. If the light is spreading outwards from a point source, the wave front at any point in space is a sphere, but if only a very small portion of the wave front is taken, at a distance of many wavelengths from the source, the wave front may be considered to be a plane surface. In fig. 3, let a , b and c be three rays forming part of a beam of

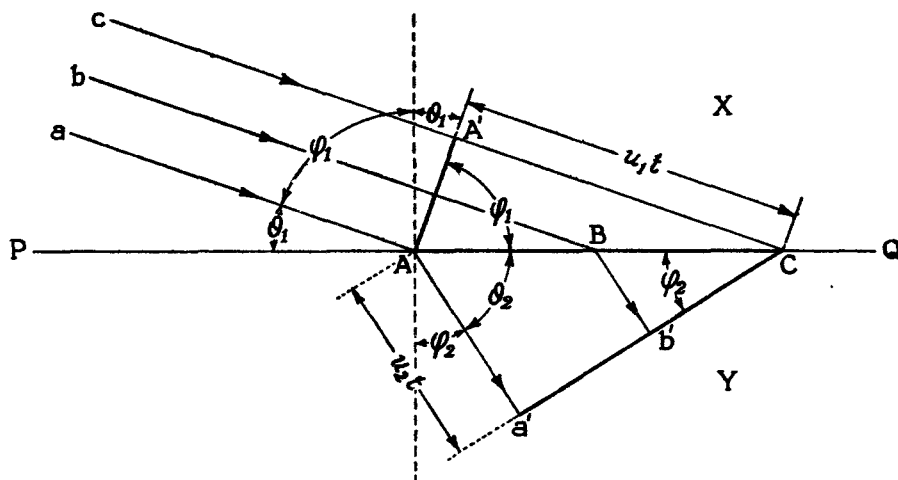


FIG. 3, CHAP. XIV.—Refraction of electro-magnetic wave.

light which is travelling with velocity u_1 in a medium X towards the surface PQ; at the instant when the light travelling along the ray a reaches the surface at A the wave front is AA' . The light travelling along the ray c must therefore pass through a distance $A'C$ before reaching the surface PQ at the point C. On reaching the surface at the point A the light of ray a enters the

medium Y, in which its velocity is u_2 . The light carried by ray C, travelling with velocity u_1 , traverses the distance A¹ C in a time $t = \frac{A'C}{u_1}$. The light travelling along the ray a, in the medium Y, will in this time move through a distance Aa' where Aa' = $u_2 t$, or

$$Aa' = u_2 \times \frac{A'C}{u_1}.$$

Hence

$$\frac{A'C}{Aa'} = \frac{u_1}{u_2}.$$

The angle of incidence is φ_1 and the angle of refraction φ_2 , by the definitions given above. Since in the diagram $\varphi_1 + \theta_1 = 90^\circ$, $\varphi_2 + \theta_2 = 90^\circ$, it is seen that

$$\frac{A'C}{AC} = \sin \varphi_1$$

$$\frac{Aa'}{AC} = \sin \varphi_2$$

$$\frac{A'C}{Aa'} = \frac{\sin \varphi_1}{\sin \varphi_2} = \nu, \text{ by definition.}$$

It follows therefore that

$$\nu = \frac{u_1}{u_2}.$$

Hence the refractive index of any two media is equal to the ratio of the velocities in the respective media. If u_1 is greater than u_2 , as in fig. 3, the refracted ray is bent towards the normal, while if u_2 is greater than u_1 it is bent in the opposite direction.

11. Although light waves have been used for illustrative purposes, the above reasoning is applicable to any form of electro-magnetic wave. It has already been stated that in a medium of permeability μ and permittivity κ the velocity of electro-magnetic waves is $\frac{c}{\sqrt{\mu\kappa}}$ centimetres per second, where c is the velocity in free space. In the case of radio waves, we are not concerned with their propagation through any medium having a permeability differing appreciably from unity, and therefore it appears that the velocity, u , should be equal to $\frac{c}{\sqrt{\kappa}}$ centimetres per second, and $\frac{c}{u} = \sqrt{\kappa}$. But $\frac{c}{u}$ is by definition the refractive index ν , and thus the refractive index of a material should be equal to the square root of its permittivity.

12. As will be shown later, the permittivity of a given material is not truly a constant, but depends to some extent upon the frequency; consequently the velocity of electro-magnetic waves, except in free space, also depends upon the frequency. White light is composed of a mixture of waves of different frequencies, and if a narrow beam of white light, that is, sunlight, is allowed to fall obliquely upon a thick sheet of plate glass, it will be found that the light is resolved into seven different colours: red, orange, yellow, green, blue, indigo, and violet. These represent groups of different frequencies, of which the lowest (red) suffers least, and the highest (violet), the greatest refraction. If it is proposed to attempt to verify the above deduction that the refractive index of a material is equal to the square root of its electrical permittivity, therefore, it is essential that both the optical and the electrical measurement should be made at the same frequency—a matter of considerable difficulty. A further complication arises, in that no substance is a perfect insulant, and the conductivity must be taken into account in making the electrical measurement. Actually the refractive index is dependent upon the degree of ionization of the medium of propagation, as will be seen later. A medium in which the velocity varies with the frequency is said to be dispersive.

CHAPTER XIV.—PARAS. 13-15

Phase and group velocity

13. In Chapter VII it was pointed out that no intelligence can be conveyed by a wave unless it is modulated in some manner. It has also been shown that a modulated wave may be regarded as the sum of a number of different components having various amplitudes and frequencies. In a dispersive medium, components of different frequencies travel with different velocities, and as a result the signal itself travels at a speed differing from that with which an unmodulated wave would be propagated in the same medium. To explain this, let us consider a transmission consisting only of two frequencies of the same amplitude. The two waves combine to form a composite group of waves, resembling in form the heterodyne beat discussed in earlier chapters, but the group itself will travel through space. Its velocity, however, will not be that of the individual waves and it is necessary to distinguish between the phase velocity and the group velocity in the particular medium. The phase velocity is that with which a single frequency of constant amplitude would be propagated, while the group velocity is that with which a signal is propagated.

14. The effect is illustrated in fig. 4, which represents the position at two successive moments

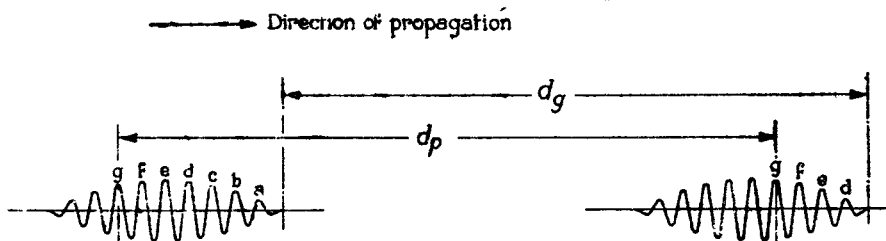


FIG. 4, CHAP. XIV.—Phase and group velocity.

of a group of electro-magnetic waves, which is moving through a cloud of electrons ; the distance travelled by the whole group during the interval is d_g and this distance, divided by the duration of the interval, is the group velocity, u_g . On the other hand we may fix our attention upon any given crest, say that marked g, and trace its progress. As the wave moves forward through the medium the vanguard a, b, c, is robbed of its energy in setting the electrons into oscillation, and this energy, less that which is dissipated by electronic collision, tends to maintain the waves as they die away. There is thus a continual eating away of the head and building up of the tail of the group of waves, with the result that the crest g gradually moves forward towards the head of the group ; the distance d_p through which this crest has travelled during the interval, divided by the interval, is the phase velocity, u_p . The phase velocity is obviously greater than the group velocity and it can be shown that, whereas u_p is greater than the velocity c of light in a vacuum, u_g is smaller than c . The phase and group velocities are in fact related by the equation $u_p u_g = c^2$.

Polarization

15. It has already been stated (Chap. VII) that electro-magnetic waves consist of a transverse vibration in space, the harmonically varying quantities being the electric and magnetic fields, which are perpendicular to each other and to the direction of propagation. The orientation of these fields is called the polarization of the wave. In the case of light radiation, e.g., by an incandescent solid, every molecule of which may be regarded as a rudimentary hertzian doublet, the vibrations take place in every conceivable plane and no definite polarization can be observed. During passage through certain media, for example crystalline quartz and Iceland spar, vibrations in other than particular planes may be suppressed, and on emergence, the wave is said to be plane-polarized. The wave emitted by a straight radiating conductor is polarized in such a manner that its electric field vector lies in a plane passing through the longitudinal axis of the conductor (Chap. VII). If the medium of propagation is free space, this state of polarization will remain unchanged. The presence of material molecules, or even of free electrons, may, however, affect the polarization of the wave during its passage.

16. In radio-communication practice, it is usual to define the polarization with reference to the surface of the earth. When the electric field vector lies in a plane perpendicular to the ground, the wave is said to be vertically polarized. It does not follow that its electric field vector is perpendicular to the ground. For example, if a wave reaches the earth from an aeroplane with an angle of incidence of 45° , and its magnetic field vector is horizontal, the electric field vector is perpendicular to the magnetic vector and also perpendicular to the direction of propagation. It is therefore in the vertical plane, but is tilted forward in the direction of propagation, making an angle of 45° with the ground. An alternative method of describing the polarization is adopted when the wave is travelling in one medium, and reaches the boundary surface of another. It is often then convenient to refer to the wave as being polarized in the plane of incidence with respect to the boundary surface, or perpendicular to this plane, as the case may be, referring to the electrical field vector in both instances. It should, however, be remembered that in most text-books dealing with the polarization of light, an older convention is adopted. A light wave is said to be polarized in the plane of incidence, when the magnetic field vector lies in this plane and the electric field vector is perpendicular thereto, which is opposite to the convention adopted in radio practice. It is easily seen that if a wave is plane-polarized at an angle with respect to the plane of incidence, the electric field vector can be resolved into two components, one in the plane of incidence and one perpendicular thereto, the two components being in phase with each other, and perpendicular to the direction of propagation.

17. If the wave front contains two mutually perpendicular electric fields of equal amplitude, but differing in phase by 90° , the wave is said to be circularly polarized. The conception is illustrated in fig. 5a, in which an electric charge Q is forced to vibrate under the action of two electric fields, one acting along the axis YOY , $\hat{r} \sin \omega t$, and one acting along the axis XOX , $\hat{r} \cos \omega t$. Under the action of the former force alone, the charge Q would execute sinusoidal

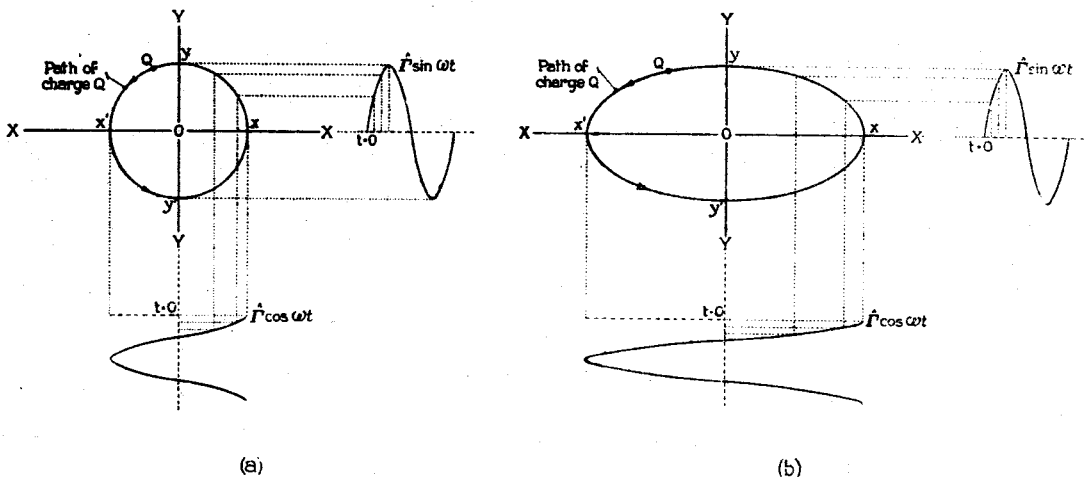


FIG. 5, CHAP. XIV.—Circular and elliptical polarization.

vibration between the points Y, Y' , while under the action of the latter alone, the charge Q would execute a sinusoidal vibration between the points X, X' . When both forces are applied simultaneously, the charge Q will travel round the circular path $X' Y' X Y$. At any instant, the charge Q must be moving in the direction of the resultant electric field, which therefore must be constantly rotating in space about the point O , with a frequency $\frac{\omega}{2\pi}$. If, however, the two fields are not of equal amplitude, although differing in phase by 90° , the conditions are those of fig. 5b.

CHAPTER XIV.—PARAS. 18-19

The charge Q will now describe in space the elliptical path shown in the diagram, and the resulting force upon it must be rotating in space while undergoing a variation of amplitude according to its orientation. Such a field is said to be elliptically polarized.

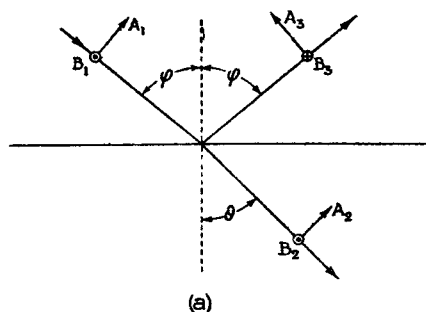
Fresnel's equations

18. The first complete theory of reflection and refraction to give results in accordance with experiment was developed by Fresnel, who was unaware of the electro-magnetic nature of light, and considered its propagation to take place in an all-pervading ether having mechanical properties resembling those of an elastic solid. Fresnel's theory was therefore based on purely mechanical phenomena, and was expressed in the form of equations giving the proportions of reflected and refracted light to that incident on the boundary surface between two media. These are known as Fresnel's equations; they are also obtainable from the electro-magnetic theory of light propounded by Maxwell, but only by the use of mathematical reasoning of an advanced character. For this reason, they are here given without proof.

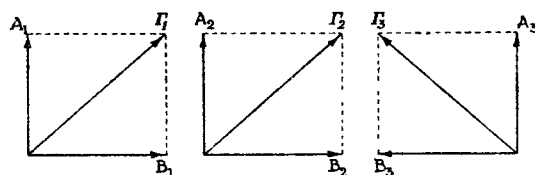
19. Suppose a parallel beam of plane-polarized electro-magnetic radiation to be incident at an angle φ upon the plane boundary surface between two non-conducting media the wave-front of the incident wave being also plane. Let the permittivity of the upper medium be κ_1 , that of the lower medium κ_2 , and the electric field strength of the incident wave be E_1 . Then E_1 may be resolved into two components, A_1 , in the plane of incidence and B_1 perpendicular thereto. Let the field strength of the refracted wave be E_2 , and its components A_2 in, and B_2 perpendicular to, the plane of refraction, which is also, of course, the plane of incidence. The angle of refraction is denoted by θ . There may also be a reflected wave of field strength E_3 , having components A_3 in, and B_3 perpendicular to, the plane of incidence. Fresnel's equations then give the values of A_2 , B_2 , A_3 , B_3 , in terms of A_1 , B_1 , and the angles of incidence and refraction. Before proceeding further, it is necessary to agree upon a convention as to the positive directions of the respective components; that usually adopted in connection with radio-communication is shown diagrammatically in fig. 6. The components in the plane of incidence, i.e., A_1 , A_2 , A_3 , are regarded as positive upwards along the surface of the paper. The components B_1 and B_2 are regarded as positive outwards from the surface of the paper, while B_3 is positive inwards from the surface of the paper. To appreciate the reason for the convention with regard to B_3 , we shall first assume its positive direction to be in the same direction as the other two. In the latter conditions, Fresnel's equations may be written

$$\left. \begin{aligned} A_2 &= A_1 \frac{2 \sqrt{\kappa_1} \cos \varphi}{\sqrt{\kappa_1} \cos \theta + \sqrt{\kappa_2} \cos \varphi} \\ B_2 &= B_1 \frac{2 \sqrt{\kappa_1} \cos \varphi}{\sqrt{\kappa_2} \cos \theta + \sqrt{\kappa_1} \cos \varphi} \end{aligned} \right\} \begin{array}{l} \text{refracted} \\ \text{wave} \end{array}$$

$$\left. \begin{aligned} A_3 &= A_1 \frac{\sqrt{\kappa_2} \cos \varphi - \sqrt{\kappa_1} \cos \theta}{\sqrt{\kappa_2} \cos \varphi + \sqrt{\kappa_1} \cos \theta} \\ B_3 &= B_1 \frac{\sqrt{\kappa_1} \cos \varphi - \sqrt{\kappa_2} \cos \theta}{\sqrt{\kappa_1} \cos \varphi + \sqrt{\kappa_2} \cos \theta} \end{aligned} \right\} \begin{array}{l} \text{reflected} \\ \text{wave} \end{array}$$



Components of electric field in plane of incidence



(b)

Positive directions, looking in direction of propagation

FIG. 6, CHAP. XIV.—Conventions used in Fresnel's equations.

20. In most practical work, the upper medium of fig. 6a is air, and κ_1 may be put equal to unity. The component B_3 then becomes

$$B_3 = B_1 \frac{\cos \varphi - \sqrt{\kappa_2} \cos \theta}{\cos \varphi + \sqrt{\kappa_2} \cos \theta}.$$

The greatest possible value of $\cos \varphi$ and $\cos \theta$ is unity. Also, if κ_2 is greater than unity, $\kappa_2 \cos \theta$ is always greater than $\cos \varphi$, and $\frac{B_3}{B_1}$ is always negative. This implies that on reflection, the component perpendicular to the plane of incidence suffers a phase change of 180° , i.e., a complete reversal of phase. By reversing the assumed positive direction as in fig. 6, then, we preserve the continuity of phase and consider instead that the sign of the amplitude is reversed. With the convention of the diagram, therefore, we must write

$$B_3 = B_1 \frac{\sqrt{\kappa_2} \cos \theta - \sqrt{\kappa_1} \cos \varphi}{\sqrt{\kappa_2} \cos \theta + \sqrt{\kappa_1} \cos \varphi}.$$

This point has been dwelt on at some length because various writers use either of the above conventions, and it is always necessary to ensure that the convention is understood, e.g., before adding the fields due to direct and reflected waves.

21. The above equations completely determine the field strength of the reflected and refracted waves, when the incident field strength and the nature of the two media are known.

Some important deductions may be made by the substitution of $\frac{\sin \varphi}{\sin \theta}$ for $\frac{\sqrt{\kappa_2}}{\sqrt{\kappa_1}}$, giving

$$A_3 = A_1 \frac{2 \sin \theta \cos \varphi}{\sin (\varphi + \theta) \cos (\varphi - \theta)}$$

$$B_3 = B_1 \frac{2 \sin \theta \cos \varphi}{\sin (\varphi + \theta)}$$

$$A_2 = A_1 \frac{\tan (\varphi - \theta)}{\tan (\varphi + \theta)}$$

$$B_2 = B_1 \frac{\sin (\varphi - \theta)}{\sin (\varphi + \theta)}.$$

The Brewster angle

22. First, if $\theta = \varphi$, both $\tan (\varphi - \theta)$ and $\sin (\varphi - \theta)$ are zero. This merely signifies that if the angle of refraction is equal to the angle of incidence, i.e., if $\kappa_2 = \kappa_1$, there can be no reflected wave, or $A_3 = B_3 = 0$. In the refracted wave $A_2 = A_1$, and $B_2 = B_1$. Second, if κ_2 is not equal to κ_1 , θ is not equal to φ and $\sin (\varphi - \theta)$ never vanishes. Also, $\sin (\varphi + \theta)$ cannot exceed

CHAPTER XIV.—PARA. 22

unity. Provided, therefore, that the component B_1 has a finite value, the component B_2 of the reflected wave will always exist. In the case of the components in the plane of incidence this is not necessarily so, for A_2 is proportional to $\frac{\tan(\varphi - \theta)}{\tan(\varphi + \theta)}$ and although the numerator is never zero, $\tan(\varphi + \theta)$ becomes infinitely great, and $A_2 = 0$, if $\varphi + \theta = \frac{\pi}{2}$. When the latter relation is fulfilled, then, the reflected wave cannot possess a component in the plane of incidence. It will be observed that if $\varphi + \theta = \frac{\pi}{2}$, $\sin \theta = \cos \varphi$ and

$$\tan \varphi = \frac{\sin \varphi}{\cos \varphi} = \frac{\sin \varphi}{\sin \theta} = \nu = \sqrt{\frac{\kappa_2}{\kappa_1}}$$

Under these conditions, the reflected and refracted components are perpendicular to each other as shown in fig. 7. The above equations embody what is known in optics as Brewster's Law, and the angle of incidence for which A_2 becomes zero is referred to as the Brewster angle φ_B . In the diagram, the upper medium is assumed to be air, $\kappa_1 = 1$, and the lower to have a permittivity $\kappa_2 = 3$. Then $\nu = \sqrt{\frac{\kappa_2}{\kappa_1}} = \sqrt{3} = \frac{0.866}{0.5} = \frac{\sin 60^\circ}{\sin 30^\circ}$, i.e., $\varphi = 60^\circ$, $\theta = 30^\circ$. The reflected ray is polarized perpendicularly to the plane of incidence, and its amplitude is equal to $B_1 \frac{\sin(\varphi - \theta)}{\sin(\varphi + \theta)} = B_1 \frac{\sin 30^\circ}{\sin 90^\circ} = 0.5 B_1$. If the incident ray contains a component A_1 , the corresponding component of the refracted wave is

$$A_2 = A_1 \frac{2 \sin \theta \cos \varphi}{\sin(\varphi + \theta) \cos(\varphi - \theta)} = \frac{2 \sin 30^\circ \cos 60^\circ}{\sin 90^\circ \cos 30^\circ} = 1;$$

thus the whole of the A component is refracted. The component B_2 is equal to

$$B_1 \frac{2 \sin 30^\circ \cos 60^\circ}{\sin 90^\circ} = 0.5 B_1,$$

i.e., the field strength of the refracted component B_2 is equal to that of the reflected component B_1 .

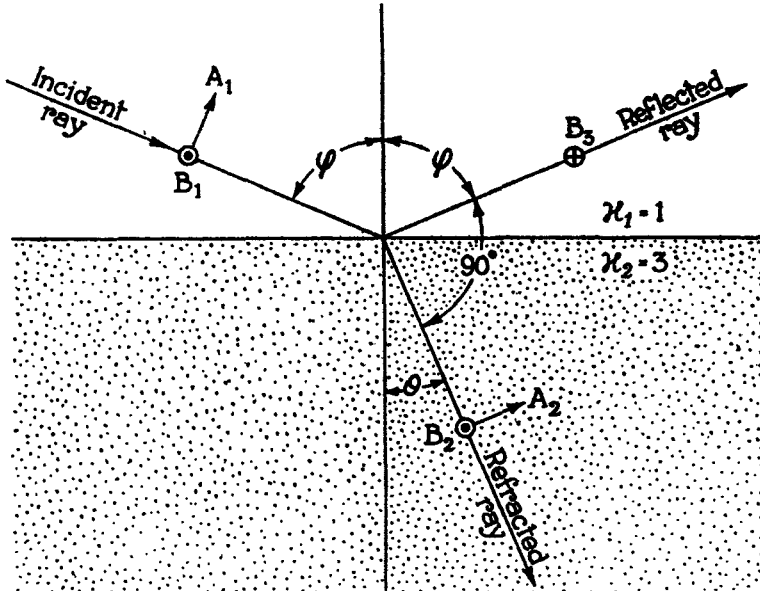
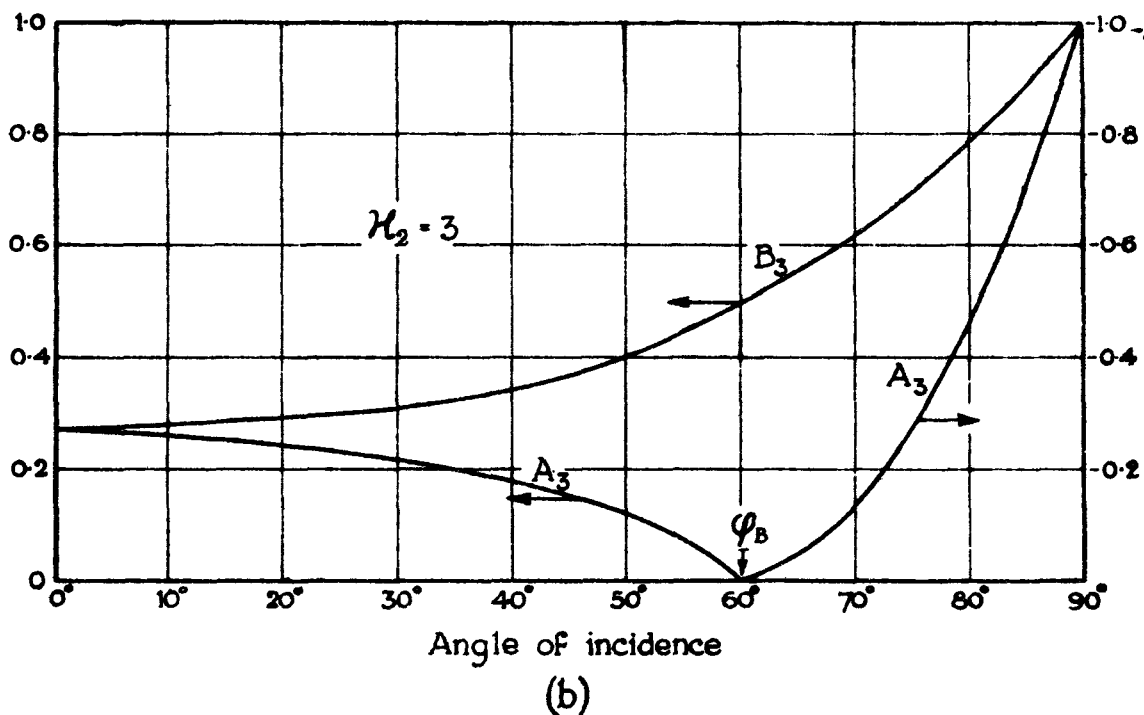
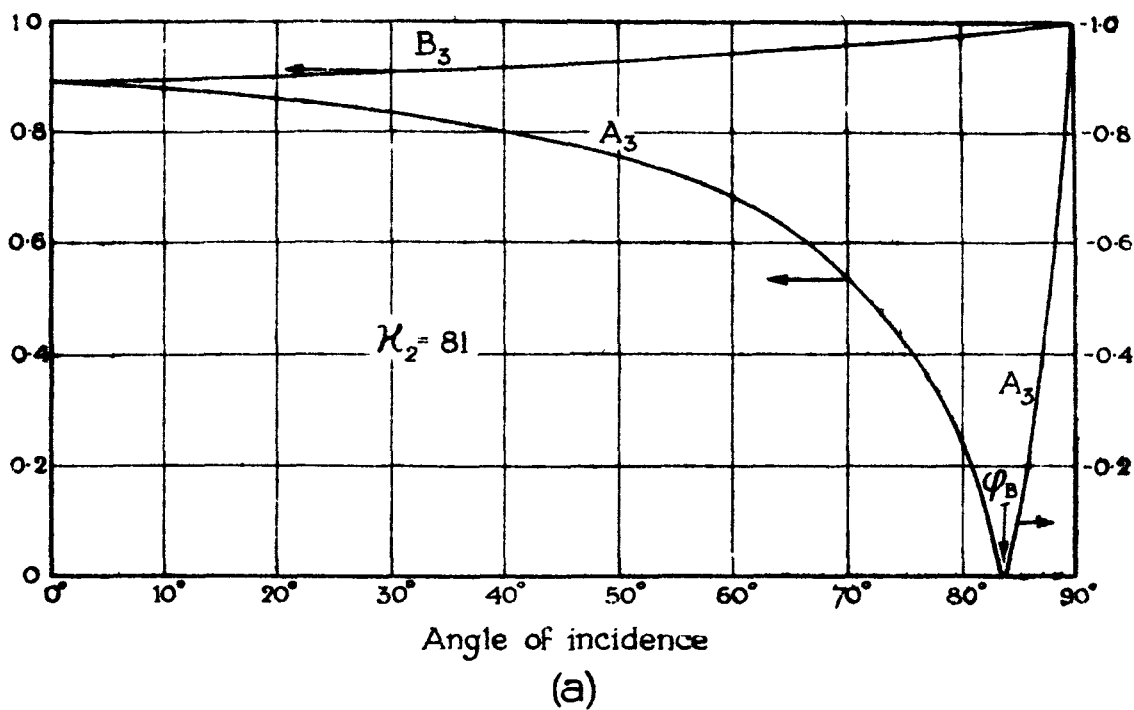


FIG. 7, CHAP. XIV.—Conditions at the Brewster angle of incidence.



AMPLITUDE OF REFLECTED WAVE

23. In radio-communication we are rarely concerned with the refracted ray, except in connection with the passage of a wave through an ionized region of the atmosphere, and in this case, since the latter has a density of ionization which varies with height, the refractive index is not constant and it is difficult to draw any useful general conclusions. Further consideration

of Fresnel's equations will therefore be confined to the reflected ray. The ratios $\frac{A_2}{A_1}$ and $\frac{B_2}{B_1}$

for various angles of incidence are shown in fig. 8a for the particular case when $\kappa_1 = 1$, $\kappa_2 = 81$; this represents the state of affairs which would arise if the second medium were pure fresh water. If the wave is incident vertically upon the surface, and $A_1 = B_1 = 1$, both A_2 and B_2 components have an amplitude of about 0.9. As φ increases the component B_1 increases gradually, becoming unity for grazing incidence. The amplitude of the component A_2 gradually decreases, becoming

zero at the Brewster angle φ_B . Since $\varphi_B = \tan^{-1} \sqrt{\frac{\kappa_2}{\kappa_1}}$ ($= \tan^{-1} 9$, in the particular instance illustrated), φ_B is seen to be 83.6° approximately. For angles of incidence greater than φ_B , the

A component suffers a phase reversal upon reflection so that the ratio $\frac{A_2}{A_1}$ is negative. Fig. 8b

shows the ratios $\frac{A_2}{A_1}$ and $\frac{B_2}{B_1}$ for $\kappa_1 = 1$, $\kappa_2 = 3$. It is seen that at vertical incidence the strength of each component is only 0.27 of the strength of the incident field.

Dispersion

24. Let us now consider the passage of an electro-magnetic wave through an ionized medium. We may obtain an idea of the phenomena involved from ordinary A.C. theory. In Chapter V it was shown that if an E.M.F. $\mathcal{E} \sin \omega t$ is applied to a circuit consisting of an inductance L and capacitance C in parallel, the supply current is

$$i_s = \left(\omega C - \frac{1}{\omega L} \right) \mathcal{E} \cos \omega t$$

If $\omega C > \frac{1}{\omega L}$, the current leads on the applied voltage by 90° , and lags by the same angle if $\frac{1}{\omega L} > \omega C$.

Now consider a condenser consisting of two parallel plates each having an area of one square centimetre, placed one centimetre apart, the space being filled by a dielectric of permittivity κ , and let a single free electron of charge e E.S. units and mass m C.G.S. units be situated in the space between the plates. If the space is traversed by the electric field of a wave of (peak) strength $\hat{F}$ E.S. units and frequency $\frac{\omega}{2\pi}$, the electron will be set in oscillatory motion at this frequency, its instantaneous velocity being

$$u_1 = \frac{e}{\omega m} \hat{F} \sin \left(\omega t - \frac{\pi}{2} \right)$$

i.e., the velocity will undergo sinusoidal variations lagging by $\frac{\pi}{2}$ radians or 90° on the variations of the electric field. It is, for the present, assumed that the motion is entirely undamped, i.e., that no energy is dissipated by the vibration of the electron. The peak value of the velocity will be

$$u_1 = \frac{e}{\omega m} \hat{F} \text{ (centimetres per second).}$$

CHAPTER XIV.—PARAS. 25-26

25. A charge e moving with a velocity of u_1 centimetres per second is equivalent to the current (E.S.U.) in an element of space one centimetre in length, or $i_1 = eu_1$. The peak value of the current will be

$$\begin{aligned} \mathcal{I}_1 &= e\mathcal{U}_1 \\ &= \frac{e^2}{\omega m} \hat{I} \\ &= \frac{\hat{I}}{\omega \frac{m}{e^2}} \end{aligned}$$

This is clearly analogous to the current flowing in an inductance L , which is given by the equation

$$\mathcal{I}_L = \frac{\mathcal{E}}{\omega L}$$

observing that in both cases the current lags on the impressed force by 90° . The free electron in the dielectric of the condenser acts as though it were an inductance

$$L' = \frac{m}{e^2} \text{ (E.S.U.)}$$

in parallel with the capacitance C , and the effective admittance of the parallel combination is

$$\omega C' = \omega C - \frac{1}{\omega L'}$$

As the condenser in this particular instance has all its geometrical dimensions equal to unity, its capacitance is $\frac{\kappa}{4\pi}$ E.S.U., i.e.,

$$\begin{aligned} \omega C' &= \omega C - \frac{1}{\omega L'} \\ &= \frac{\omega \kappa}{4\pi} - \frac{e^2}{\omega m} \\ \therefore C' &= \frac{\kappa}{4\pi} - \frac{e^2}{\omega^2 m} \\ &= \frac{1}{4\pi} \left(\kappa - \frac{4\pi e^2}{\omega^2 m} \right) \end{aligned}$$

26. (i) The effective permittivity of the dielectric is therefore not κ but $\kappa - \frac{4\pi e^2}{\omega^2 m}$; if instead of a single electron we consider the effect of N electrons in the unit volume of dielectric, the current will be Neu_1 E.S. units and the effective dielectric constant becomes

$$\kappa' = \kappa - \frac{4\pi Ne^2}{\omega^2 m}$$

Since the permittivity of a vacuum is unity, the above consideration shows that the permittivity of a vacuous space containing N electrons per cubic centimetre is $1 - \frac{4\pi Ne^2}{\omega^2 m}$. The phase velocity,

u_p , of propagation of electro-magnetic waves in entirely empty space being $\frac{c}{\sqrt{\kappa}}$ centimetres per second, it follows that, in space occupied only by free electrons

$$\begin{aligned} u_p &= \frac{c}{\sqrt{\kappa'}} \\ &= \frac{c}{\sqrt{1 - \frac{4\pi Ne^2}{\omega^2 m}}}, \end{aligned}$$

i.e., the phase velocity is greater (and the group velocity less) than the velocity c in free space

(ii) If positive ions (i.e., gaseous atoms minus one or more electrons) are also present, these will also affect the value of κ' and therefore the phase and group velocities, but since the mass of such an atom is at least 1840 times as great as that of the electron the effect on the permittivity is very small. In the above expression N represents the number of free electrons per cubic centimetre since $e = 4.77 \times 10^{-10}$ E.S. units and $m = 8.8 \times 10^{-28}$ gram,

$$\begin{aligned}\frac{e^2}{m} &= \frac{4.77 \times 10^{-10}}{8.8 \times 10^{-28}} \\ &= 2.6 \times 10^8\end{aligned}$$

It is convenient to refer to the frequency f rather than the angular velocity ω of the wave, giving

$$\begin{aligned}u_p &= \frac{c}{\sqrt{1 - \frac{4\pi N \times 2.6 \times 10^8}{4\pi^2 f^2}}} \\ &= \frac{c}{\sqrt{1 - 8.27 \times 10^7 \frac{N}{f^2}}},\end{aligned}$$

thus the phase velocity of the wave in an ionized medium depends upon the frequency; the higher the frequency, the smaller is the effect of the ionization, and we realize why the refractive index of a medium, as measured by observations at very high frequency, i.e., visible light waves, may differ from that obtained by measurement of the permittivity at radio-communication frequencies. Since u_p is greater than c , a wave, in crossing the boundary between free space and an ionized medium, will be refracted away from the normal.

Effect of ionization gradient

27. Now let us suppose that in a certain region the ionization increases progressively. To take a concrete case, let the line A C in fig. 9 represent the lower boundary of a medium in which the number of electrons per cubic centimetre increases uniformly in the upward direction. Let a ray P A enter the medium with a velocity c , at an angle θ_1 . Just inside the boundary, if the refractive index is v , the phase velocity will be

$$u_p = \frac{c}{v},$$

and $v \sin \theta_r = \sin \theta_1 = \text{constant}$.

The ray is therefore bent at the point of entry; during its passage through a further small element of distance in the medium, the value of v changes, but so also does the angle of the ray with reference to the boundary surface. At any point where the direction of propagation makes an angle θ with the normal to the boundary, $v \sin \theta$ remains constant. It follows that the effect of the increasing electronic density, or reduction of refractive index, will be to increase the angle θ , which may eventually become 90° . The ray then commences to pass through successive strata of progressively decreasing ionization, and is therefore bent downwards until it finally passes through the boundary at an angle equal to θ_r , with which it entered. At the highest point B, where the electron density is, say, N_B electrons per cubic centimetre, $\theta_B = 90^\circ$, $\sin \theta_B = 1$, and

$$\begin{aligned}v_B &= \sin \theta_1 = \frac{c}{u_p} \\ &= \sqrt{1 - \frac{8.27 \times 10^7 N_B}{f^2}} \\ \text{or } v_B &= 0 \text{ if } N_B = \frac{f^2}{8.27 \times 10^7}.\end{aligned}$$

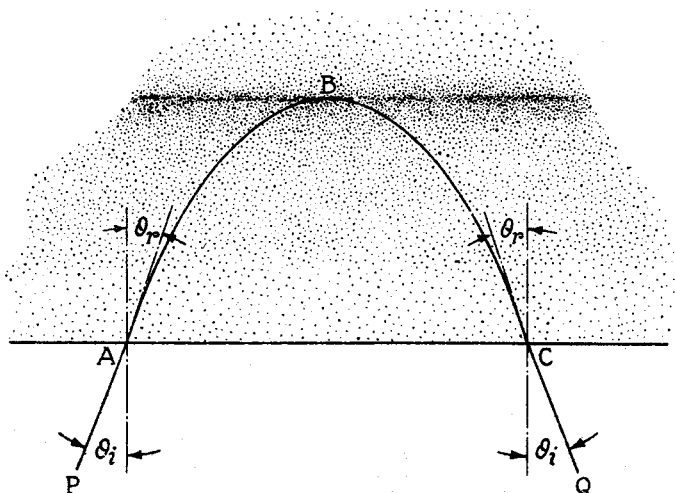


FIG. 9, CHAP. XIV.—Refraction in an ionized region.

If the electron density exceeds this figure, the direction of travel of the ray will be completely reversed. It will be noted that a region having a refractive index of zero is impenetrable by an electro-magnetic wave.

Relation between ionization and conductivity

28. If the wave is travelling through a space containing matter, e.g., an ionized gas, collisions occur between electrons (or ions) and the gas molecules, and it can be shown that if p such collisions take place per second, the permittivity of the medium is

$$\kappa' = \kappa - \frac{4\pi Ne^2}{(p^2 + \omega^2)m}$$

The effect of these collisions is to cause a dissipation of energy, and we may infer that the collision frequency p will enter into any expression giving the specific resistance, or alternatively, the conductivity, of the medium. The conductivity of a gaseous medium is

$$\sigma = \frac{pNe^2}{(p^2 + \omega^2)m} \text{ (E.S.U.)}$$

The change of permittivity and the conductivity of the medium are evidently closely related physically. As already shown, the reduction in permittivity is due to a velocity component 90° out of phase with the field, while the conductivity is represented by a velocity component in anti-phase with the field causing the motion. The passage of a wave through a material of finite conductivity therefore takes place with a progressive decrease of amplitude, and the wave is said to be attenuated. The conductivity is sometimes expressed in E.S.U., sometimes in E.M.U. and sometimes as the reciprocal of the specific resistance ρ (ohms per centimetre cube). The following conversion factors are therefore given.

$$\begin{aligned} \sigma \text{ (E.S.U.)} &= \sigma \text{ (E.M.U.)} \times (9 \times 10^{10}) \\ &= \sigma' \times 9 \times 10^{11} \\ \text{where } \sigma' &= \frac{1}{\rho} \end{aligned}$$

Effect of conductivity

29. (i) We may obtain some idea of the effect of conductivity in a medium in which an electro-magnetic wave is propagated, by first considering the motion of an electron which is situated between the parallel plates of a condenser of unit dimensions, the permittivity of its

dielectric being κ' . We may assume that this dielectric has a number of free electrons, the conductivity being a measure of the ability or otherwise of these electrons to execute sinusoidal vibration in the presence of an electric field. The effect of a lack of freedom is to reduce the amplitude of vibration and to cause the velocity of the electrons to lag by less than 90° on the electric field producing the motion.

(ii) Let the conductivity of the medium be σ E.S. units and a sinusoidal P.D. of v E.S. units be set up between the plates of the condenser. As a result of this P.D. an alternating current will flow. This current will consist of a displacement component

$$i_d = j\omega C v$$

and a conduction component

$$i_c = g v$$

so that the total current is (in E.S.U.)

$$i = (g + j\omega C) v.$$

Since the condenser is of unit dimensions, $C = \frac{\kappa'}{4\pi}$ and $g = \sigma$, hence

$$\begin{aligned} i &= \left(\sigma + j\omega \frac{\kappa'}{4\pi} \right) v \\ &= \frac{j\omega}{4\pi} \left(\kappa' - j \frac{4\pi}{\omega} \sigma \right) v \\ &= \frac{j\omega}{4\pi} \left(\kappa' - j \frac{2\sigma}{f} \right) v. \end{aligned}$$

30. (i) In effect, therefore, the permittivity of the conductive medium is $\kappa' - j \frac{2\sigma}{f}$ instead of κ as in a perfect dielectric. It is now necessary to apply this reasoning to the special case where the electric field between the plates is set up by an electro-magnetic wave, which is propagated parallel to them in such a manner that the electric field vector of the wave is in the same direction as the P.D. just discussed. It must be observed that since the only object of the plates is to distribute the charge over a given volume of dielectric, we may now ignore their existence, thus removing complications due to the presence of the highly conductive boundary surfaces.

(ii) The equation of a travelling wave of frequency $\frac{\omega}{2\pi}$ is given in Chapter VII. In the present instance we may write for the electric field

$$v = \hat{r} \cos \frac{2\pi}{\lambda} (u_p t - x),$$

where u_p is the phase velocity. As already shown, in a perfect dielectric $u_p = \frac{c}{\sqrt{\kappa}}$. Eliminating λ , then,

$$v = \hat{r} \cos \omega \left(t - \frac{x}{u_p} \right)$$

and for a conductive dielectric, instead of $u_p = \frac{c}{\sqrt{\kappa}}$, we must write $u_p = \frac{c}{\sqrt{\kappa' - j \frac{2\sigma}{f}}}$ so that

$$v = \hat{r} \cos \omega \left(t - \frac{x}{c} \sqrt{\kappa' - j \frac{2\sigma}{f}} \right).$$

To interpret this equation, we use Demoivre's theorem (Chapter V). It is there shown that $\cos \theta$ is the real part of $e^{j\theta}$. Since the real and imaginary parts of a complex number are entirely independent we may perform any operation on $e^{j\theta}$, and take the real part of the result to give

CHAPTER XIV.—PARA. 31

the effect of the operation upon $\cos \theta$. In the present example then, if

$$\begin{aligned} \gamma &= \hat{r} \cos \omega \left(t - \frac{x}{c} \sqrt{\kappa' - j \frac{2\sigma}{f}} \right) \\ &= \text{real part of } \hat{r} e^{j\omega \left(t - \frac{x}{c} \sqrt{\kappa' - j \frac{2\sigma}{f}} \right)} \\ \text{and } \sqrt{\kappa' - j \frac{2\sigma}{f}} &= \nu - j\alpha, \\ \gamma &= \text{real part of } \hat{r} e^{j\omega \left(t - \frac{x}{c} \nu + j \frac{x}{c} \alpha \right)} \\ &= \text{real part of } \hat{r} e^{j\omega \left(t - \frac{x}{c} \nu \right)} e^{-\frac{\omega x}{c} \alpha} \\ &= e^{-\frac{\omega x \alpha}{c}} \hat{r} \cos \omega \left(t - \frac{x}{c} \nu \right). \end{aligned}$$

Hence the phase velocity is $\frac{c}{\nu}$. The exponential factor $e^{-\frac{\omega x \alpha}{c}} = e^{-\frac{2\pi \alpha}{\lambda} x}$ gives the rate at which the amplitude decreases as the wave progresses.

31. If both κ' and $\frac{\sigma}{f}$ are known it is easy to find numerical values for ν and α . Since

$$\begin{aligned} \sqrt{\kappa' - j \frac{2\sigma}{f}} &= \nu - j\alpha, \\ \kappa' - j \frac{2\sigma}{f} &= \nu^2 - 2j\nu\alpha - \alpha^2 \\ \therefore \kappa' &= \nu^2 - \alpha^2 \\ \frac{2\sigma}{f} &= 2\nu\alpha \\ (\kappa')^2 &= \nu^4 - 2\nu^2\alpha^2 + \alpha^4 \\ \left(\frac{2\sigma}{f}\right)^2 &= 4\nu^2\alpha^2 \\ (\kappa')^2 + \left(\frac{2\sigma}{f}\right)^2 &= \nu^4 + 2\nu^2\alpha^2 + \alpha^4 \\ \sqrt{(\kappa')^2 + \left(\frac{2\sigma}{f}\right)^2} &= \nu^2 + \alpha^2 \end{aligned}$$

Hence

$$2\nu^2 = \sqrt{(\kappa')^2 + \left(\frac{2\sigma}{f}\right)^2} + \kappa'$$

$$2\alpha^2 = \sqrt{(\kappa')^2 + \left(\frac{2\sigma}{f}\right)^2} - \kappa'$$

i.e.,

$$\nu = \sqrt{\sqrt{\left(\frac{\kappa'}{2}\right)^2 + \left(\frac{\sigma}{f}\right)^2} + \frac{\kappa'}{2}}$$

$$\alpha = \sqrt{\sqrt{\left(\frac{\kappa'}{2}\right)^2 + \left(\frac{\sigma}{f}\right)^2} - \frac{\kappa'}{2}}.$$

These formulæ have an important bearing upon the depth of penetration of a radio wave into the earth. If κ' is small compared with $\frac{2\sigma}{f}$, we may write

$$\alpha = \sqrt{\frac{\sigma}{f} - \frac{\kappa'}{2}}$$

Thus if $\sigma = 10^8$ E.S.U., $f = 1$ Mc/s, $\kappa' = 10$, $\frac{\sigma}{f} = 100$ and

$$\begin{aligned}\alpha &= \sqrt{95} \\ &= 9.75.\end{aligned}$$

The rate at which the amplitude of the wave is reduced is given by

$$I_x = I_0 e^{-\frac{\omega x}{c}}$$

The amplitude will be reduced to $\frac{1}{e} = 0.367$ of its original value in passing through a distance x , where

$$\begin{aligned}\frac{\omega x}{c} &= 1 \\ x &= \frac{c}{\omega \alpha} \\ &= \frac{3 \times 10^{10}}{2\pi \times 10^6 \times 9.75} \\ &= 440 \text{ centimetres.}\end{aligned}$$

Also $v = \sqrt{105} = 10.25$. The phase velocity in the ground is less than one-tenth of that in free space.

Effect of magnetic field

32. When the ionized region through which an electro-magnetic wave is propagated is also occupied by a magnetic field, the character of the wave may be considerably modified during its passage. An attempt will be made to show the manner in which these effects are produced. We may first consider a charge of Q (E.M.U.) moving with constant velocity U through a uniform

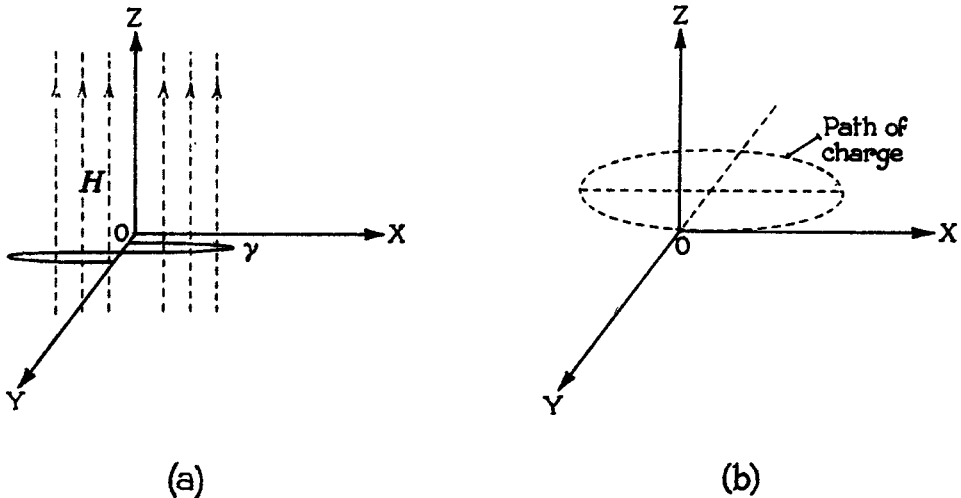


FIG. 10, CHAP. XIV.—Effect of magnetic field.

CHAPTER XIV.—PARA. 32

magnetic field of strength H (E.M.U.). Since a moving charge is equivalent to an electric current, there will be a force F acting upon the charge, its direction being perpendicular to the directions both of H and U , and its magnitude will be $F = Q H U$ dynes. Let the charge be carried by a mass of m grams, and at a time $t = 0$ to be situated at the origin O of the rectangular co-ordinates $O X$, $O Y$, $O Z$, where H is along $O Z$, as in fig. 10a. Suppose a stationary sinusoidal electric field $\gamma = \hat{r} \sin \omega t$ to be applied along the axis $O X$. At $t = 0$, let the instantaneous velocity of the charge along $O X$ be $u_x = \frac{dx}{dt}$ and along $O Y$ be $u_y = \frac{dy}{dt}$. Then the charge will be subject to forces F_x and F_y , where

$$F_x = Q (\hat{r} \sin \omega t - H u_y)$$

$$F_y = Q (H u_x).$$

Since force = mass $\times$ acceleration, we may write

$$m \frac{d^2 x}{dt^2} = Q \left(\hat{r} \sin \omega t - H \frac{dy}{dt} \right) \quad \dots \dots \dots (a)$$

$$m \frac{d^2 y}{dt^2} = Q H \frac{dx}{dt} \quad \dots \dots \dots (b)$$

From (b)

$$m \frac{d^2 y}{dt^2} = Q H \frac{d^2 x}{dt^2}$$

$$\frac{d^2 x}{dt^2} = \frac{m}{QH} \frac{d^2 y}{dt^2}$$

inserting this value for $\frac{d^2 x}{dt^2}$ in (a)

$$\frac{m^2}{QH} \frac{d^2 y}{dt^2} + QH \frac{dy}{dt} = Q \hat{r} \sin \omega t$$

$$m \frac{d^2 y}{dt^2} + \frac{Q^2 H^2}{m} \frac{dy}{dt} = \frac{Q^2 H}{m} \hat{r} \sin \omega t$$

This equation can be immediately integrated, giving

$$m \frac{d^2 y}{dt^2} + \frac{Q^2 H^2}{m} y = - \frac{Q^2 H}{\omega m} \hat{r} \cos \omega t + K$$

The constant K depends on the initial conditions and may be made zero by a suitable choice of these. The solution of an equation of the form

$$P \frac{d^2 y}{dt^2} + Ry = S \cos \omega t$$

is given in standard text-books on differential equations. The complete solution will consist of a forced, i.e., sustained oscillation and a damped oscillation. The forced oscillation along $O Y$ will take place in accordance with the equation

$$y = \frac{Q^2 H}{\omega (\omega^2 m^2 - Q^2 H^2)} \hat{r} \cos \omega t,$$

and the motion along the axis $O X$ is given by

$$x = \frac{Qm}{\omega^2 m^2 - Q^2 H^2} \hat{r} \sin \omega t.$$

Comparing these results with those of para. 17 it will be seen that the path of the charge will be an ellipse, as shown in fig. 10b.

33. (i) The point of immediate interest in the above expressions is that, theoretically, the amplitudes of vibration along both O X and O Y become infinite if $\omega^2 m^2 = Q^2 H^2$ or

$$\omega = \frac{QH}{m}.$$

Actually, of course, the amplitude would gradually build up from zero and would take an infinite time to reach an infinite amplitude. The position is in fact analogous to that reached in considering the current in an acceptor circuit with zero resistance. The presence of the magnetic field introduces into the motion of the charge a kind of resonance effect. If the charge, hitherto denoted by Q (E.M.U.) is that of an electron, the resonant frequency is $\frac{\omega}{2\pi} = \frac{eH}{mc}$ where e is the charge of an electron in E.S.U.

(ii) When the sinusoidal electric field, instead of being stationary, is due to the propagation of an electro-magnetic wave, the effects are somewhat complicated, depending upon the direction of the magnetic field relative to the direction of propagation and to the polarization of the wave on entering the field. Assuming the wave to be plane polarized on entry, if the electric field vector E is parallel to the magnetic field vector H , the latter has no effect upon the propagation characteristics. If, however, E is perpendicular to H , various effects are produced, depending upon the direction of propagation. If the latter is parallel to H , the plane polarized wave is split into two circularly polarized components, which undergo different degrees of absorption and refraction. If both the direction of propagation and the electric field vector are perpendicular to H , e.g., if the former is along O X and the latter along O Y, it is split into two plane-polarized waves, which again undergo unequal degrees of absorption and refraction. In general, the propagation is along neither of the axes, but in some intermediate direction, and the emergent wave is elliptically polarized.

Fresnel's equations for conductive media

34. Fresnel's equations may be applied to the problem of finding the respective amplitudes of the reflected and refracted waves at the boundary surface between two conductive media, by substituting the complex permittivity of each medium for the simple permittivity appropriate to a perfectly non-conducting substance. If the incident wave is in air or free space, its permittivity is unity and its conductivity zero. Let the permittivity of the second medium be κ , and its conductivity σ . Its complex permittivity is then $\kappa' = \kappa - j\frac{2\sigma}{f}$, where f is the frequency of the incident disturbance. Then

$$A_3 = A_1 \frac{\kappa' \cos \varphi - \sqrt{\kappa' - \sin^2 \varphi}}{\kappa' \cos \varphi + \sqrt{\kappa' - \sin^2 \varphi}},$$

$$B_3 = B_1 \frac{\sqrt{\kappa' - \sin^2 \varphi} - \cos \varphi}{\sqrt{\kappa' - \sin^2 \varphi} + \cos \varphi},$$

the positive directions being as shown in fig. 6. Analogous expressions for the refracted wave are easily deduced from the equations of paragraph 19 if required.

Reflection coefficient, horizontal polarization

35. When the reflection takes place at the surface of the earth, A_3 is identified as the vertically polarized component and B_3 as the horizontally polarized component of the field strength in the reflected wave. It is now convenient to change the notation slightly. The ratios $\frac{A_3}{A_1}, \frac{B_3}{B_1}$, may be termed the reflection coefficients for the respective states of polarization and are complex quantities. The reflection of a horizontally polarized wave will be first dealt with, as the results

CHAPTER XIV.—PARA. 36

are somewhat the simpler. The ratio $\frac{B_2}{B_1}$, is a complex number and may be written

$$\frac{B_2}{B_1} = K_h / \theta_h = \frac{\sqrt{\kappa - j \frac{2\sigma}{f} - \sin^2 \varphi - \cos \varphi}}{\sqrt{\kappa - j \frac{2\sigma}{f} - \sin^2 \varphi + \cos \varphi}}$$

Then K_h is the arithmetical ratio of the strengths of the reflected and incident fields and θ_h is an angle which must be added to the phase of the incident beam to obtain the phase of the reflected wave. The positive directions must be taken as in fig. 6 when making such additions. The angle θ_h is always negative, lying between zero and -180° . If K_h and θ_h are calculated for given values of the constants κ , σ , f , and plotted against the angle of incidence φ , the curves obtained always resemble those marked K_h and θ_h in fig. 11a and fig. 11b. respectively.

36. The numeric K_h is very easily calculated for the particular case when $\varphi = 0$, corresponding to vertical incidence, as in the following example.

$$\text{Let } \kappa = 9, \frac{2\sigma}{f} = 10, \varphi = 0$$

$$\cos \varphi = 1, \sin \varphi = 0$$

$$K_h / \theta_h = \frac{\sqrt{9 - j10} - 1}{\sqrt{9 - j10} + 1}$$

$$\text{Let } \sqrt{9 - j10} = v - j\alpha$$

$$9 - j10 = v^2 - 2j\alpha v - \alpha^2$$

$$v^2 - \alpha^2 = 9$$

$$2\alpha v = 10$$

$$(v^2 - \alpha^2)^2 = v^4 - 2v^2\alpha^2 + \alpha^4 = 81$$

$$(2\alpha v)^2 = 4v^2\alpha^2 = 100$$

$$\therefore v^4 + 2v^2\alpha^2 + \alpha^4 = 181$$

$$\text{and } v^2 + \alpha^2 = \sqrt{181} \\ = 13.44$$

$$v^2 - \alpha^2 = 9$$

$$\therefore 2v^2 = 22.44$$

$$v = 3.34$$

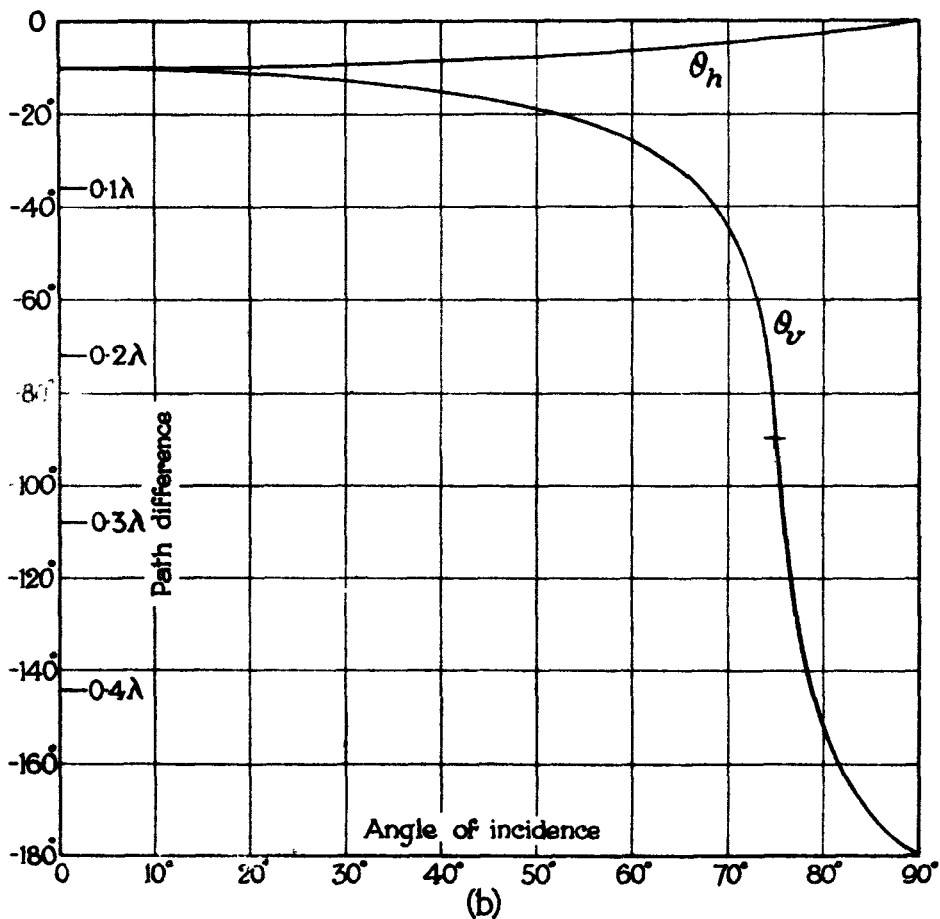
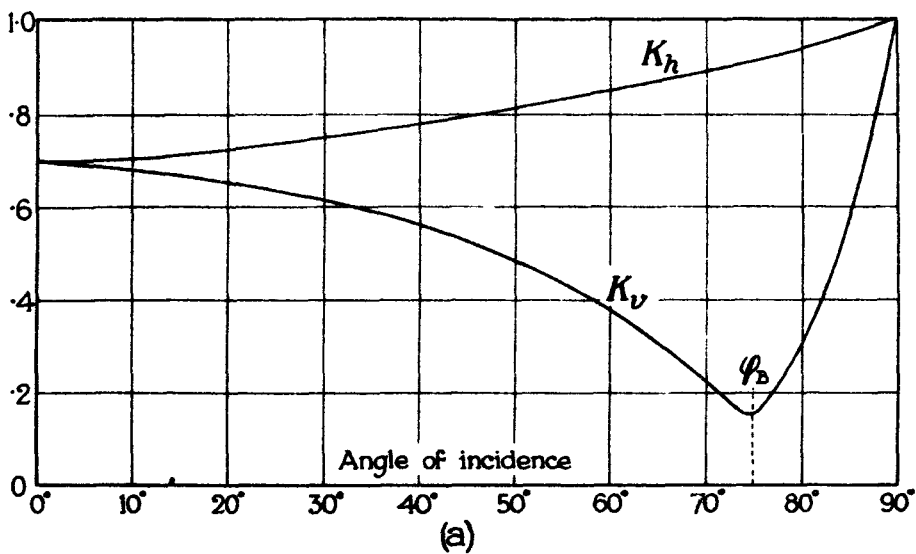
$$\text{and } 2\alpha^2 = 4.44$$

$$\alpha = 1.49$$

$$\text{Therefore } \sqrt{9 - j10} = 3.34 - j1.49$$

The accuracy of this result may be checked by squaring $(3.34 - j1.49)$

$$(3.34 - j1.49)^2 = 11.17 - 2j4.98 - 2.22 \\ = 8.95 - j9.96.$$



REFLECTION COEFFICIENTS
 $(\kappa = 9, \frac{2\sigma}{f} = 10)$

FIG. II
 CHAP. XIV

We now have

$$\begin{aligned} K_h / \theta_h &= \frac{3.34 - j 1.49 - 1}{3.34 - j 1.49 + 1} \\ &= \frac{2.34 - j 1.49}{4.34 - j 1.49} \\ &= \frac{13.7 - j 3}{21} \end{aligned}$$

$$K_h = 0.67$$

$$\begin{aligned} \theta_h &= \tan^{-1} \left(-\frac{3}{13.7} \right) \\ &= -\tan^{-1} 0.22 \\ &= -12\frac{1}{2}^\circ \text{ approximately.} \end{aligned}$$

When the wave is incident at an angle of 90° , $\cos \varphi = 0$, and therefore $K_h = 1$, $\theta_h = 0^\circ$. The complete variation of K_h and θ_h with φ , for $\kappa = 9$, $\frac{2\sigma}{f} = 10$, is shown in fig. 11.

37. When a radio receiver is subject to the action of a downcoming wave, the beam is so wide that the incident and reflected waves cannot be separated as they can be when a parallel beam of light is reflected from a plane surface. The receiver is in general subject to both the incident and reflected beam and it is necessary to investigate the resultant beam when both are present. In fig. 12 let us consider the field at a point P, situated at a height h above the ground. The electric field due to the incident beam along D P sets up a field $\gamma = \hat{r} \cos \omega t$, at the point P, measured positively above the surface of the paper. The reflected wave arrives at P by the path A B P. Draw H B perpendicular to D P and let H P = d . Then at the point of incidence B the incident wave has a field strength $\hat{r} \cos \left(\omega t + \frac{2\pi d_1}{\lambda} \right)$ measured positively above the surface of the paper. The wave reflected at the point B produces at P the field

$$K_h \hat{r} \cos \left\{ \omega t + \frac{2\pi}{\lambda} (d_1 - d_2) + \theta_h \right\}$$

which is positive in the direction below the surface of the paper. Consideration of the figure shows that $d_2 - d_1 = 2h \cos \varphi$, and the electric field at P is therefore

$$\hat{r} \left\{ \cos \omega t - K_h \cos \left[\omega t + \theta_h - \frac{2\pi}{\lambda} (2h \cos \varphi) \right] \right\}$$

38. This expression has an interesting application in the reception on the ground of horizontally polarized waves emitted by an aircraft. If the aircraft is very remote from the receiver, φ will be nearly 90° , and if h is small, $\frac{2\pi}{\lambda} (2h \cos \varphi)$ will also be a very small angle. When φ approaches 90° , K_h is nearly unity and θ_h only a few degrees, hence the electric field strength at the point P is the resultant of two fields of almost equal intensity which are very nearly in antiphase, and is thus very small. If, however, the receiving aerial is at a height h , where

$$\frac{2\pi}{\lambda} (2h \cos \varphi) = \pi, \text{ or } h = \frac{\lambda}{4 \cos \varphi},$$

the two fields are of almost equal intensity and arrive at P practically in phase, giving a very strong signal.

CHAPTER XIV.—PARAS. 39-40

Reflection coefficient, vertical polarization

39. Turning now to a consideration of the vertically polarized wave, the phenomena are seen to be more complicated than in the former instance. The ratio of reflected to incident field strength is

$$\frac{A_3}{A_1} = K_v / \theta_v = \frac{\left(\kappa - j \frac{2\sigma}{f}\right) \cos \varphi - \sqrt{\kappa - j \frac{2\sigma}{f} - \sin^2 \varphi}}{\left(\kappa - j \frac{2\sigma}{f}\right) \cos \varphi + \sqrt{\kappa - j \frac{2\sigma}{f} - \sin^2 \varphi}}$$

Here again K_v is the numerical value of the ratio and θ_v is an angle to be added to the phase of the reflected wave; θ_v is always negative, lying between zero and -180° . It is easily seen, by putting $\varphi = 0$, $\cos \varphi = 1$, $\sin \varphi = 0$, that for vertical incidence, K_v / θ_v is equal to K_h / θ_h . As the angle of incidence increases, K_v decreases and reaches a minimum value. Over this range of variation of φ , θ_v approaches the value -90° , at first slowly but afterwards more rapidly.

The value of φ at which $\theta_v = -\frac{\pi}{2}$ is known as the pseudo-Brewster angle; instead of complete extinction, such as would occur at the surface of a perfect dielectric, a marked reduction in amplitude of the reflected wave occurs when the wave is incident at the pseudo-Brewster angle. For greater values of φ , the change of phase increases very rapidly and is -180° when $\varphi = 0$.

The variation of K_v and θ_v , with φ , is also shown in fig. 11, for $\kappa = 9$, $\frac{2\sigma}{f} = 10$.

40. The field due to a downcoming wave and its reflection can be worked out in a manner similar to the previous case, but it is not so easy to draw any definite conclusions as to the effect of elevation of the receiving aerial. Reference to fig. 12 shows that if at the point P there is a

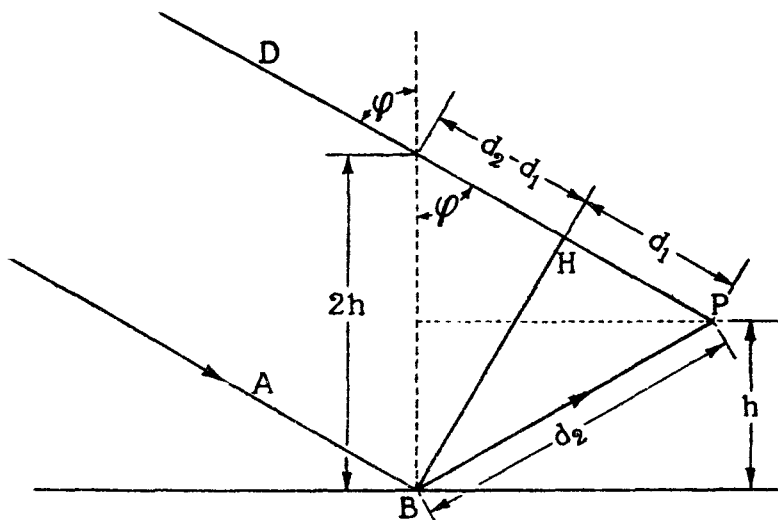


Fig. 12, CHAP. XIV.—Effect of direct and reflected waves upon elevated aerial.

field $\hat{F} \cos \omega t$ due to the direct action of the downcoming signal, there will be a reflected field of strength

$$K_v \hat{F} \cos \left[\omega t + \theta_v - \frac{2\pi}{\lambda} (2h \cos \varphi) \right]$$

If ϕ is very nearly 90° and greater than the pseudo-Brewster angle, while h is small, K_r will be nearly unity and θ_r nearly -180° , so that the two fields tend to annul each other. Near the ground, therefore, an almost grazing beam and its reflection nearly cancel each other in their effect upon a receiving aerial.

Propagation along ground

41. At first sight, the results obtained in paras. 37, 38 and 40 lead to the conclusion that neither a horizontally nor a vertically polarized wave can be propagated along the earth's surface, for at grazing incidence the direct and reflected waves completely cancel each other. It must however be remembered that the theory is not completely in accordance with fact. Both the incident wave-front and the reflecting surface are assumed to be plane, whereas plane waves are physically unrealizable and the earth's surface is also curved. The theory must therefore not be regarded as an attempt to deny the experimentally established fact that the radiation from a transmitting aerial may be propagated along the earth's surface to some extent, but merely as showing that the mechanism of the ground wave, as this radiation is called, is more complex than it appears. The exact mode of propagation is still debatable. It appears certain that the phenomenon of diffraction is responsible for the propagation to a limited extent, but diffraction alone will not account for the propagation of waves of frequencies higher than about 500 kc/s. According to certain physicists, a vertical aerial on or near the ground gives rise to what is called a surface wave, in addition to the radiation predicted by the simple theory given in Chap. XV. This surface wave must not be confused with the ground wave referred to above, which is assumed to be part of the radiation predicted by the simple theory, but affected by the ground constants in such a manner that it is guided round the protuberance of the earth. The surface wave, where it exists, is strongest at the surface, the field strength both above and below becoming rapidly weaker. Experiments at different frequencies in some cases appear to support and in others to deny the existence of the vertically polarized surface wave. Both theory and experiment, however, agree that a horizontally polarized wave cannot be propagated along a conductive surface to an appreciable distance. Suppose a wave to be initiated in such a manner that it starts to travel over and parallel to the earth's surface, plane polarized at an angle of 45° . The electric field then has components A_1 in the vertical and B_1 in the horizontal plane. Assuming for the moment that A_1 remains vertical, B_1 will be reduced to zero after travelling a comparatively short distance, for it induces radio-frequency currents in the earth, and the energy so dissipated must be supplied by the energy originally possessed by the component B_1 .

42. The following account of an experiment actually performed many years ago at the National Physical Laboratory, Teddington, may be of interest. In order to verify or disprove the above conclusion, an attempt was made to produce a wave which travelled over the surface of the earth with horizontal polarization, the transmitting "aerial" being simply a large solenoid of 42 turns, 5 feet in diameter, which was placed with its axis in the vertical plane. This was connected to a condenser having negligible external field, thus constituting a closed, feebly radiating oscillatory circuit. The polarization of the radiated wave at a receiver only a few miles away was found to be vertical. A second experiment for the same purpose produced exactly the same effect. In this case the aerial of a transmitting station was completely removed; the counterpoise (which measured 600 feet by 150 feet, and was 9 feet above the ground) was cut into two equal portions which were then energized as the aerial and counterpoise respectively of a transmitting aerial system. The natural polarization of such a radiator is horizontal, and it was thought possible that at Slough, a distance of a few miles, the horizontally polarized component of the wave would be at least as strong as the vertically polarized component, but the energy actually received was found to be almost entirely due to a vertically polarized wave.

43. An additional effect of the finite conductivity of the earth is to cause the wave front to be tilted forward in the direction of propagation, because the velocity just within and just above the ground must be the same, i.e., the velocity cannot change abruptly. The velocity at ground level is therefore a few parts in one hundred less than the velocity at a considerable height. Suppose the waves to be polarized in the plane of incidence, and the angle of tilt at the surface of

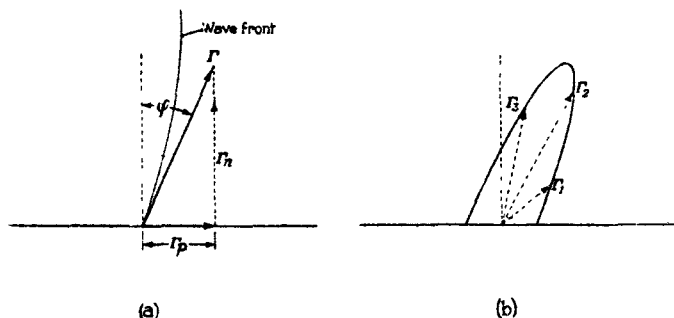


FIG. 13, CHAP. XIV.—Tilting of wave at earth's surface.

the earth (measured from the normal) to be ϕ (fig. 13a). The electric field vector then consists of two components, (i) a truly vertical component E_v and (ii) a horizontal component E_h , having its positive direction in the direction of propagation. This component sets up eddy currents in the ground, and as a result of the energy dissipation the horizontal and vertical components are not exactly in phase. The end of a vector representing the instantaneous strength of the electric field will therefore travel round on the elliptical path as shown in fig. 13b. For given values of α and σ , the tilt angle increases with frequency, so that a high-frequency wave suffers greater absorption than a low-frequency wave.

Attenuation

44. From the foregoing considerations, it will be readily understood that after its inception, the subsequent history of a radio wave largely depends upon the properties of the various media through (and over) which it must travel. These properties affect the direction of the propagation, the strength of the received signal and the polarization of the wave. As the wave travels outward from its point of origin, the strength of the field decreases and the wave is said to be attenuated. The attenuation may be considered as due to the three following causes.

(i) The natural spreading of the wave. This is of the same form, geometrically, as the spreading of an electric field from a point of charge; if a transmitting aerial were situated in free space, the wave front would be a spherical surface, and the energy density at any point distant x units of length from the radiator would be proportional to $\frac{1}{x^2}$. The strength of the field is proportional to the square root of the energy density and it therefore follows that the field strength, at a distance from the radiator, would be proportional to $\frac{1}{x}$. The reduction of field strength with distance may be referred to as the natural or geometrical attenuation.

(ii) The energy abstracted from the wave due to its passage through any medium possessing finite conductivity. This may be compared with the absorption of light. A piece of window glass may pass only 90 per cent. of the light reaching it, a portion of the remainder being reflected and a portion expended in heating the glass. A second, similar, piece of glass will pass only 90 per cent. of that reaching it, so that the two together will pass only 81 per cent. and so on. This absorption follows the law of logarithmic decay, and may be referred to as logarithmic or exponential attenuation.

(iii) Finally, since no radio transmitter is situated in free space, but is either upon or comparatively near to the surface of the earth, attenuation is caused by the passage of the wave over the earth's surface, owing to the finite conductivity of the latter.

Early propagation formulæ

45. Considering the radiation from a vertical aerial of effective height h_e metres situated upon a perfectly conductive earth and carrying a uniform current of I amperes (R.M.S.) at a frequency of f cycles per second, the field strength at a distance of r kilometres, so small that the logarithmic attenuation is negligible, is given by the expression $\Gamma = \frac{4\pi}{10^4} \frac{h_e f I}{r}$ (microvolts per metre). With the frequencies in use in the earliest history of radio-communication, it was found that this formula gave inaccurate results at distances greater than a few miles. It will be noticed that according to the above expression the field strength should increase with frequency, whereas it was found that actually this was not the case, lower frequencies giving a better signalling range for a given value of f . After considerable experimental investigation a factor of the form $\epsilon - A r \sqrt{f}$ was added in order to correct for the logarithmic attenuation, and the formula became

$$\Gamma = \frac{4\pi}{10^4} \frac{h_e f I}{r} \epsilon - A r \sqrt{f},$$

various values of the constant A being proposed by various workers.

46. Before dismissing these formulæ it may be pointed out that whereas the first factor is derived theoretically on the assumption that the waves are propagated outwards in such a manner that the wave front at any instant is a hemispherical shell, the second factor was deduced empirically. Prior to about 1922 an expression of this type was generally used to forecast the performance of a given aerial; it will be seen that since the factor $\epsilon - A r \sqrt{f}$ increases as f decreases, it appears that better signalling ranges should be obtained with low than with high radio frequencies. As a result of this deduction, long-distance communication was performed almost entirely on frequencies below 60 kc/s, some stations of very high power being operated on frequencies as low as 12 kc/s. Frequencies of the order of 1,000 kc/s and above were almost entirely neglected, or in certain instances deliberately adopted to restrict the signalling range for purely local working.

47. In this connection, however, it is unfair to be severely critical of the research workers of this period, for although it was recognized as early as 1912 that the future progress of radio-communication would depend upon the development of C.W. signalling and heterodyne reception, it was not until much later that the valve oscillator was sufficiently developed to be of value for long-distance transmission. Prior to this the C.W. generators available (such as the radio-frequency alternator and Poulsen arc) were either suitable only for low frequency (L/F and VL/F) working, or capable of operation at a much higher efficiency under these conditions. The employment of the lower radio frequencies for signalling purposes was therefore dictated as much by the generators available as by the supposedly superior propagation of these waves. When suitable C.W. oscillators became available, shortly after the end of the war of 1914-1918, systematic research work into the propagation of short waves was undertaken by service, commercial and amateur workers. The great value of the amateur contribution to this research was chiefly due to their world-wide distribution and elaborate though entirely voluntary organization for reporting and verifying the reception of signals transmitted by other amateurs. It was found that the foregoing propagation formulæ were of little value for frequencies higher than about 1,000 kc/s, the obtainable ranges being sometimes very much greater than predicted thereby. The mode of propagation of these waves was therefore seen to be more complex than had hitherto been thought.

48. As long ago as 1882, before the production of electro-magnetic waves by the direct action of electric currents had been achieved, it had been suggested by Balfour Stewart, as a result of the study of terrestrial magnetism, that at a height of about 200 kilometres, instead of being an almost perfect insulant as it is at ground level, the earth's atmosphere should have a conductivity of the same order as that of the earth itself. When, in 1907, Marconi succeeded in signalling between England and Newfoundland, this suggestion was recalled almost simultaneously by Heaviside in England and Kennelly in America. Both these scientists pointed out that if such a conductive region actually existed, wireless waves originating on the earth would be

CHAPTER XIV.—PARA. 49

unable to penetrate it, but would be reflected downwards and so would travel round the protuberance of the earth. Kennelly in particular stated clearly that the effect of such a stratum in the atmosphere would be to decrease the geometrical attenuation ; instead of the field strength varying inversely as the distance r , it would vary inversely as the square root of the distance. Accordingly, it was proposed to predict the probable range of a given transmitter by means of a formula of the form

$$\Gamma = \frac{4\pi h_e f I}{10^4 \sqrt{r}} - 4r\sqrt{f}$$

but this was also found to fail when applied to the propagation of high frequencies.

Nature of the earth's atmosphere

49. Before proceeding further, it is necessary to discuss briefly the nature of the atmosphere. This consists chiefly of nitrogen, water vapour and oxygen, together with traces of many other

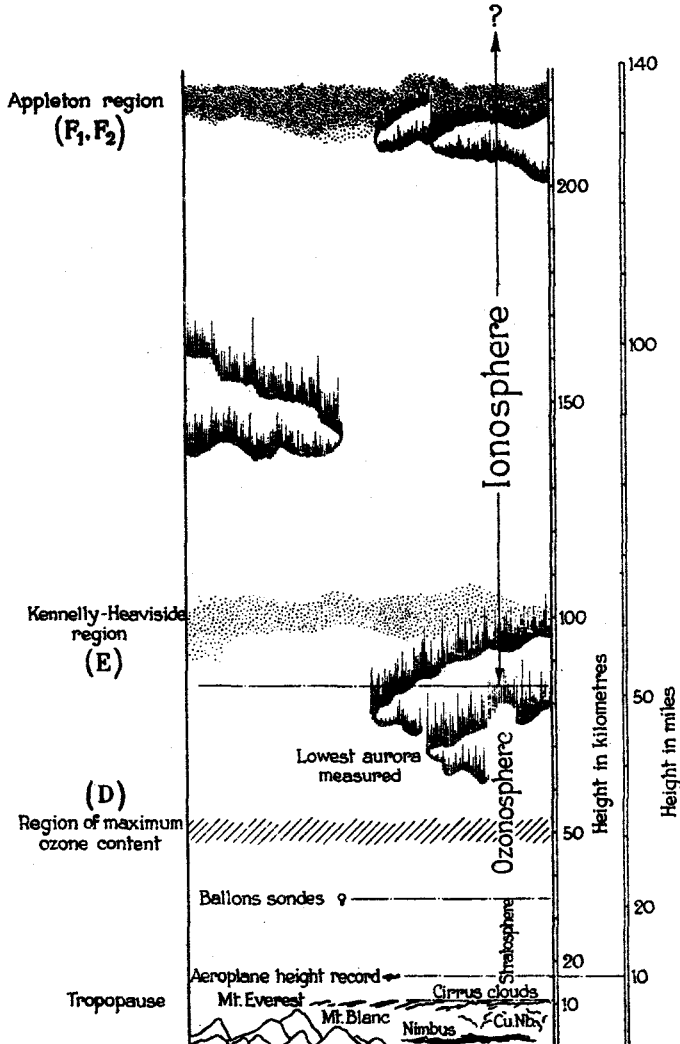


FIG. 14, CHAP. XIV.—The earth's atmosphere.

gases, and is divided into several regions which are shown pictorially in fig. 14. The region immediately above the earth's surface is called the troposphere. Its chief characteristic is the turbulent motion of the air. Winds, cloud formation, rain, etc., are all found in this strata. In the temperate zones the "roof" of the troposphere is about 10 kilometres above ground, and is called the tropopause. In the troposphere the temperature decreases with height, an average of about 6°C . per kilometre. Thus the temperature at the tropopause is about -50°C . over Europe. Over the equator its height is rather more—about 15 kilometres—and its temperature about -75°C .

50. Above the tropopause we have a region of about 20 kilometres deep, called the stratosphere. Human exploration of this region has so far been confined to its lowest levels, but a great deal of knowledge regarding it has been obtained by recording instruments carried by free balloons and by observations of the behaviour of meteors. At about 20 kilometres above ground the pressure is roughly one-twentieth that at the surface and the temperature over the temperate zones about -55°C ., while it is lower still over the equator. The upper portion of the stratosphere is also called the ozonosphere, because a comparatively small quantity of ozone is distributed throughout its volume, over a height of from 40 to 60 kilometres. Although small in quantity this ozone absorbs a very large proportion of the ultra-violet radiation from the sun. The temperature of the ozonosphere is raised to some 40°C . as the result of this absorption of solar energy. It is probable that at such heights both oxygen and nitrogen are also present.

The ionosphere

51. The name "stratosphere" was originally given because it was formerly thought that in this region there is little or no wind, and the various gases therefore tend to separate into strata, the heaviest being of course nearest the earth. According to this theory there was little difference between night and day in this region, its temperature being nearly constant and of the order of -50°C . It is now known that the atmosphere at high levels is warmed during the day and cooled at night, so that winds are set up, at any rate up to heights of the order of 100 kilometres. The temperature in the higher portions of the stratosphere may in fact vary between $+100^{\circ}\text{C}$. and -50°C . at different times of the day and year. The exact agency by which portions of this region become ionized is still in some doubt, but it is almost certain that the most important factor is the emission of ultra-violet rays by the sun. Under certain conditions, it is possible that particles of electrically neutral matter are shot out by the sun, and these may also play a part. At present, however, it is considered that the sun's ultra-violet radiation is sufficiently intense to account for the observed phenomena. The ionized region is usually referred to as the ionosphere. Assuming that the ionizing agent is ultra-violet radiation, it will be understood from fig. 15 that

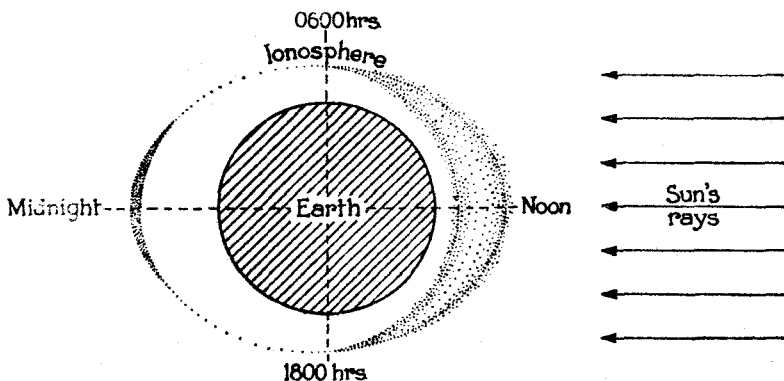


FIG. 15, CHAP. XIV.—Formation of ionized regions.

when any particular portion of the ionosphere is within the shadow of the earth, that portion receives little or no ionizing radiation. The positive and negative ions (or free electrons) into which

CHAPTER XIV.—PARAS. 52-55

it is separated during daylight, will therefore tend to re-combine during the dark hours. During daylight over any particular part of the earth then, the ionized region is of greater depth and its lower boundary (which is of course very ill-defined) is nearer the earth than during the night.

The ionized regions E and F

52. Direct evidence of the existence of conducting layers in the upper regions of the atmosphere was first obtained by Appleton and Barnett in 1925. In brief, the later experiments of this series consisted of the transmission of a very short train of waves—called a pulse—which was received at a point near the transmitter. As a rule it was found that on the transmission of a single pulse, three or more pulses were received, the first arriving at the receiver practically instantaneously with the transmission, owing to the proximity of the two stations. The other pulses were received at distinct intervals after the first, and in all cases the interval was consistent with the assumption that the wave had travelled upward until it reached a reflecting layer and had then returned. The first set of measurements, using a frequency of 750 kc/s, indicated that the reflector was situated about 90 to 100 kilometres above the surface of the earth. This phenomenon was expected, and the height at which the reflection occurred was in accordance with theoretical estimates. What the experimenters had achieved, at this stage, was to establish the existence and approximate height of an ionized region.

53. Research on these lines has now proceeded for a number of years and, as a result, we are able to speak with confidence of two regions of maximum ionic density. For brevity these regions are sometimes referred to as "layers," but it must be understood that they possess no definite boundaries. It is convenient to refer to the virtual height of a layer as the height at which signals appear to commence their return journey to earth, or at which the apparent reflection takes place. The critical penetration frequency of any region is defined as the lowest radio-frequency which will penetrate that region at normal (i.e., vertical) incidence. We may think of the two regions as concentric shells of ionized gas surrounding the earth, the inner, which is known as the "E" or Kennelly-Heaviside layer being at a distance of about 100 kilometres above the earth, and the outer, which is known as the "F" or Appleton layer, at an average height of about 230 kilometres. Under certain conditions each of these may be split into two regions of maximum ionic density. We then speak of the E_1 and E_2 , F_1 and F_2 layers, E_2 being slightly higher than E_1 and F_2 slightly higher than F_1 . In all cases, the virtual heights are referred to.

54. The mechanism by which ionization is set up is thought to be somewhat as follows. Assuming that the principal cause of ionization is the ultra-violet radiation of the sun, we have to consider the absorption of energy which ionization implies. At very great heights there is very little gas present so that although the radiation is intense, very few ions are formed, and very little energy is lost by the radiation. As the latter approaches the earth the gas pressure increases and ionization commences, but since the mere fact of ionization causes loss of energy, the intensity of the radiation falls off rapidly as it approaches the earth. We therefore find the state of affairs to be somewhat as shown by the solid curve of fig. 16, which is drawn on the assumption that the ionizing radiation has a single frequency and that only one kind of gas molecule is present. Actually, of course, this is not the case, and there may be several regions of maximum ionization. In any case the ionizing influence exerted at a lower level, e.g., in the E region, will depend upon the amount of solar radiation absorbed at higher levels.

55. As previously suggested, ions and electrons in a dissociated state have a strong tendency to combine. In the ionosphere, the rate of re-combination depends mainly upon the gas pressure. This is of course much higher at 100 kilometres than at 230 kilometres and consequently the rate of re-combination is much higher in the E than in the F layer. In addition to the E and F regions, it appears probable that portions of the ozonosphere also become ionized, although to a considerably smaller degree than the upper regions. The existence of this region—which has been allotted the letter D—may account in part for the propagation of a ground wave at the lower radio-frequencies, and its virtual disappearance at the higher frequencies. It is probable that the D layer ionization is due in part to the bombardment of gas molecules by cosmic rays.

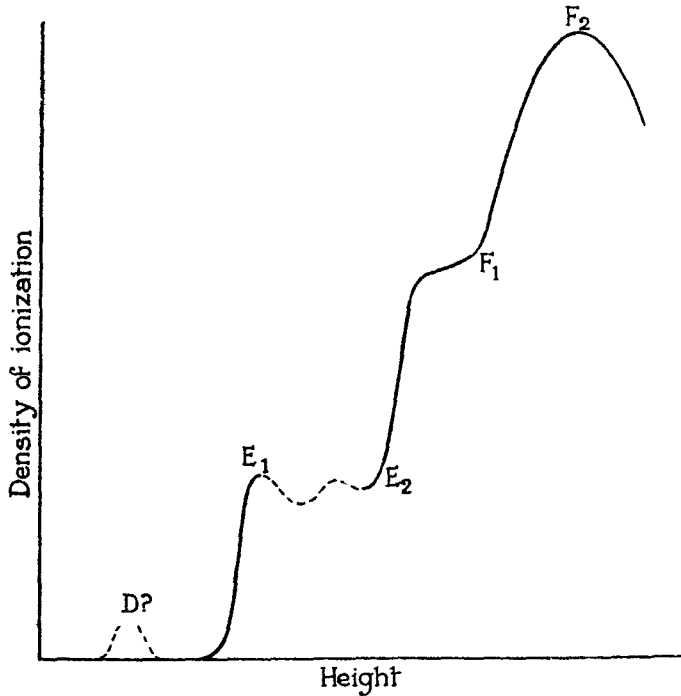


FIG. 16, CHAP. XIV.—Variation of ionization with height.

Seasonal, geographic and solar influences

56. The virtual height of an ionized layer is an important characteristic in that it determines the maximum frequency at which waves will be reflected. For a given transmission distance the angle of incidence at a layer will be greater, the lower the layer. The critical frequency for a given ionization density varies approximately as the secant of the angle of incidence, so that for a given density and signalling distance, a lower layer will reflect waves of a higher frequency. Thus, even though the ionization density of the E layer is considerably lower than in the F layer, the former may on occasion determine the maximum frequency which may be usefully employed. The state of the ionosphere also varies with the latitude; in general, the density of ionization is greatest in equatorial regions. It also varies seasonally, being greatest in the summer and least in winter. Because the seasons are less clearly defined in equatorial regions than in others, the seasonal variations are of less practical importance than in northern or southern latitudes. In addition to variations associated with day and night conditions, latitude and seasons, progressive changes in the ionosphere occur during the period of the sun-spot cycle—approximately eleven years. The ionization density is highest when the sun-spot activity is a maximum; it is now increasing and may be expected to reach a maximum in the year 1939. Observations have shown that certain solar disturbances are associated with pronounced variations which tend to repeat at intervals of 27·3 days, this being the period of rotation of the sun. Whether the solar disturbance and the ionization variations are related in the manner of cause and effect, or whether both are due to some unknown cosmic phenomenon, is at present an open question.

Change of polarization

57. Referring to paragraph 33, it will be appreciated that the earth's magnetic field will have a considerable influence upon the passage of a wave through the ionosphere. During its

CHAPTER XIV.—PARAS 58-59

passage, the wave may be split into two circularly or elliptically polarized components, one having a right-handed and the other a left-handed sense of rotation, looking in the direction of propagation. The former is referred to as the extraordinary and the latter as the ordinary ray. When such splitting occurs, it is necessary to distinguish between the critical frequencies of the two components, and the following notation is used. The symbol f_c is used to denote the critical frequency in general. The critical frequency for the ordinary ray in the F_1 region is denoted by $f_{o_{F_1}}$, and for the extraordinary ray $f_{x_{F_1}}$. Similarly in the F_2 region the critical frequencies are denoted by $f_{o_{F_2}}$ and $f_{x_{F_2}}$. The critical frequency for the E region is denoted by f_E , the electrical and magnetic characteristics at this height being such that double refraction (i.e., splitting) occurs only infrequently. In the northern hemisphere the right-handed component, i.e., the extraordinary ray, is absorbed to a greater degree than the left-handed component, and *vice versa* in the southern hemisphere. The greater part of the absorption appears to take place in the lower regions, where the atmospheric pressure is fairly high, the collision frequency large and the ionic density small.

58. Both the absorption and the polarization depend upon the magnitude of $\frac{eH}{mc}$ where e is the charge of an electron in E.S.U., m its mass in grams and $c = 3 \times 10^{10}$. The strength H of the earth's magnetic field varies from about 0.35 oersted at the equator to about 0.6 oersted at the poles. By making certain simplifying assumptions it is possible to calculate, approximately the frequency f_r at which the electronic oscillations exhibit quasi-resonant properties. Thus if $H = 0.4$ oersted

$$\begin{aligned} f_r &= \frac{1}{2\pi} \frac{eH}{mc} \\ &= \frac{4.77 \times 10^{-10} \times 0.4}{6.28 \times 8.8 \times 10^{-28} \times 3 \times 10^{10}} \\ &= 1.15 \times 10^6 \text{ cycles per second.} \end{aligned}$$

At this frequency, a plane-polarized wave having its electric field vector perpendicular to H , and propagated along the axis OY in fig. 10a, would set the electrons in the ionized region into violent oscillation and numerous collisions would occur between electrons and gas molecules. This would give rise to severe absorption and consequently the frequencies in the quasi-resonant band are but poorly propagated *via* the ionosphere. According to the direction of propagation with respect to the magnetic field, and the latitude, it is found that the frequency band 1 to 2 Mc/s (approximately) suffers almost complete absorption owing to this phenomenon, so that communication by this band must depend chiefly upon the ground wave.

Fading

59. Fading is observed to some extent at all frequencies, but is more pronounced at high frequencies than at low. The term fading is generally applied to variations of signal strength, occurring over only a short interval of time, e.g., a few minutes, but an allied phenomenon is the complete cessation of communication on a given frequency for many hours or even days, after which communication is restored. The latter form of fading often occurs when the transmission path passes near the magnetic poles, and thus is believed to be connected with the mechanism of terrestrial magnetism. The latter phenomenon in turn is associated with the sun-spot activity; it has been observed that when a sun-spot passes over the sun's meridian, as viewed from the earth, a period of fading is generally experienced at a later epoch, which may be from one to three days. Rapid fading, on the other hand, is generally attributed to interference between waves arriving at the receiver by paths of different lengths, and by a variation of the nature of polarization. The effect of interference is similar to that caused by reflection at the surface of the earth (paras. 34-40), but is complicated by the fact that the ionization along the path of the wave through the ionosphere is subject to slow fluctuations, while the respective planes of polarization of the ground and sky waves may also change. If a vertical aerial is in use at a

ground station, the strength of the received signal depends only upon the strength of the vertically polarized component, and if the degree of vertical polarization changes, the signal strength will fluctuate accordingly.

60. About three years ago it was noticed that at certain times, long distance H/F communication would completely fail for periods of 15-30 minutes, sometimes 45 minutes, over the whole of the surface of the earth that was illuminated at the time by the sun. It was found by astronomers that this fading occurred simultaneously with "bright eruptions" of hydrogen gas from the surface of the sun in the neighbourhood of a sun-spot. These bright eruptions produce an intense ultra-violet radiation which is believed to cause a great increase in the intensity of ionization in the E layer. It is probable that the increased attenuation in the E layer is the cause of this kind of fading. During such a period, the transmission on L/F and VL/F is sometimes improved. It was thought at first that this type of fading is subject to a 54-day periodicity, but recent evidence has not confirmed this. It may be expected that H/F fading will occur more frequently and increase in intensity until the sun-spot maximum (1939). It will probably decrease to a minimum about 1944 and so on with the eleven year sun-spot period.

Propagation of low-frequency waves (L/F and VL/F)

61. When the frequency of the signal is below about 300 kc/s, and the range of reception does not exceed about 1,000 miles, most of the energy reaches the receiver by the ground wave, and the season and time of day have little effect. On the other hand, at ranges of from 2,000 to 4,000 miles or so, the energy received is practically confined to that carried by the sky wave, and the strength of signals shows pronounced diurnal and seasonal variations owing to the changes in location and density of the E layer. Between 1,000 and 2,000 miles, both the sky and ground waves are of appreciable strength. Both these are emitted in the same phase, but they do not traverse paths of the same length, and do not, except fortuitously, arrive at the receiver in phase; the field at the receiving aerial is the vector sum of the two, and if the effective length of either path varies, the signal strength will also vary. When the whole of the transmission path is in daylight, the signal strength of a very distant transmitter is relatively low, but more or less constant, but as the sunset line falls across the transmission path, the field strength generally falls; when however the whole path is in darkness the intensity usually increases to a high level, which is maintained until sunrise.

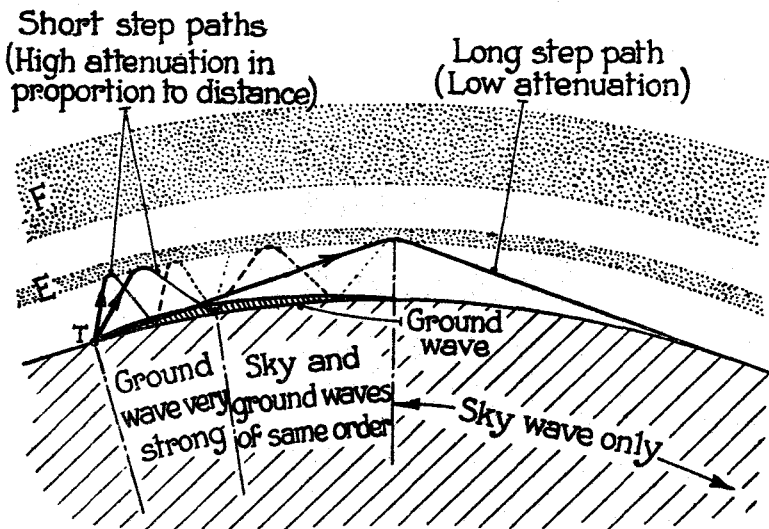


FIG. 17, CHAP. XIV.—Propagation of low and medium frequencies.

CHAPTER XIV.—PARAS. 62-66

62. The propagation of these waves is illustrated in fig. 17 which shows the route taken by several hypothetical rays, according to the vertical angle of radiation at the transmitter. At such low frequencies, the E layer is highly refracting, and the upward-travelling waves are bent downward towards the earth after relatively little penetration, that is, the bending process resembles the reflection of a light wave. The wave, on reaching the earth, is again reflected, and travels upward once more, much energy being lost by absorption at each reflection. The energy which travels in the space between the layer and earth suffers only geometrical attenuation, and consequently, the greater part of the energy received at extreme ranges is that which leaves the transmitter at a low vertical angle and is reflected at the layer only once. It is not, however, possible to design a low-frequency aerial in such a manner as to concentrate the energy in the desired direction, for this would necessitate an aerial of vertical height equal to several half-wavelengths.

63. The diurnal and seasonal variations in signal strength result from variations in the height and density of the E layer ; when the layer is low, or the electron density high, the attenuation will be high. As the frequency is increased, these variations become more marked, until, at frequencies of the order of 500 kc/s, long-distance transmission is not possible during daylight and is subject to enormous fluctuation during dark hours. As already observed, low-frequency waves travelling near the surface of the earth are always normally polarized, and the wave front is tilted forward slightly as a result of the absorption of energy by the conductive earth. Fading is rarely observed during daylight but may be troublesome when signalling over extreme ranges at night.

The 600-1,500 kc/s band

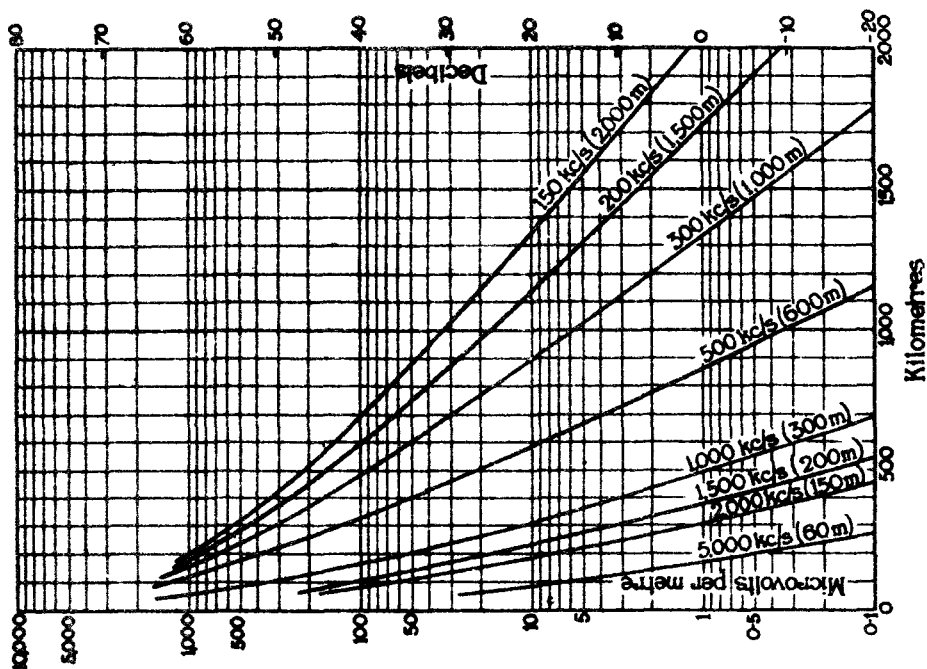
64. These frequencies are chiefly used for the broadcasting of entertainment programmes, and their propagation has been studied by a very large number of investigators. The chief point to be observed is that with this type of transmission, what is required is a comparatively loud, and very constant, signal over a region within a few hundred miles of the transmitter. The region in which reasonably high-quality telephony can be received is that in which the ground wave is powerful enough to give a high degree of discrimination against all interference. The field strength required in industrial districts is much higher than in rural ones, owing to the higher noise level. Experience indicates that for satisfactory service a field strength of from 5 to 30 millivolts per metre is necessary in towns, whereas a field of only 1 millivolt per metre may suffice in the country.

Field-strength curves (ground wave)

65. Fig. 18a shows the variation with distance in the intensity of the ground wave, when the transmission is wholly over sea water. The left-hand scale gives the field strength for a radiated power of one kilowatt ; if P kilowatts are radiated the field-strength scale must be multiplied by $\sqrt{P}$. The right-hand scale gives the field intensity in decibels with reference to an arbitrary level of 1 micro-volt per metre. If P kilowatts are radiated the quantity $10 \log_{10} P$ decibels must be added to this scale. For example, at a frequency of 1,000 kc/s the field strength at a distance of 900 kilometres is 10 micro-volts per metre, which is 20 db. above the reference level. A radiated power of 4 kilowatts gives a field strength of 20 micro-volts per metre, which is $(20 + 10 \log_{10} 4) = 26$ db. above the reference level. Fig. 18b is similar to fig. 18a except that transmission is assumed to take place over soil of average conductivity. The field strengths actually obtained may be greater or less than those given by the curves, owing to the possible presence of the sky wave as explained above.

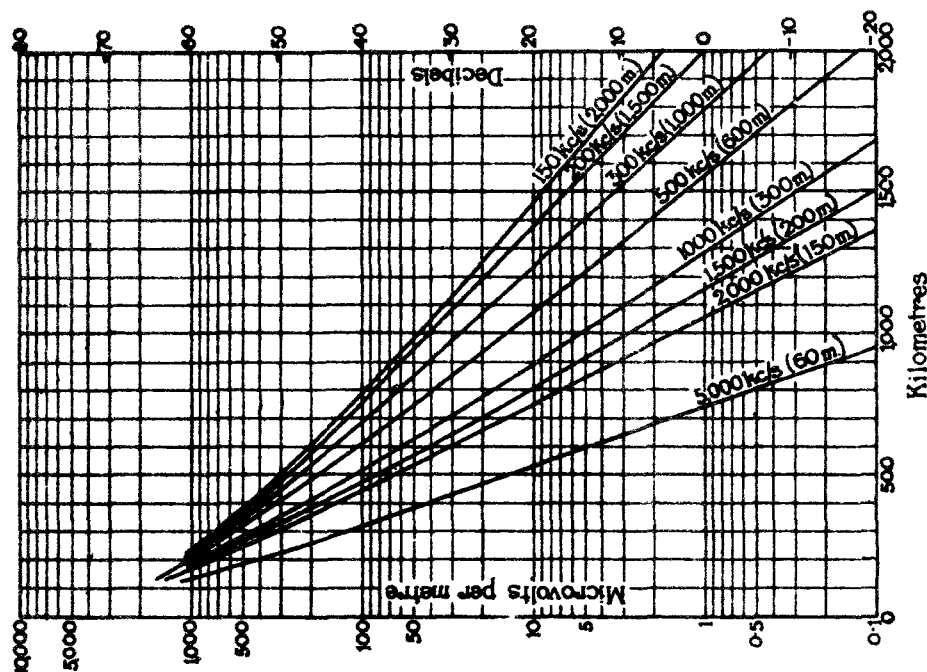
Field-strength curves (night propagation)

66. When the whole of the transmission path is in darkness, the curves given in fig. 19 may be used to estimate the probable intensity at the receiver, for all frequencies between 150 and 1,500 kc/s. The two sets of curves correspond with those of fig. 18 (for sea water and average soil respectively). The quantity plotted is referred to as the quasi-maximum field strength (per kilowatt radiated). This is defined as the strength which is exceeded by the actual field strength for only five per cent. of the time. The average field strength is, however,



AVERAGE SOIL

$$\sigma = 9 \times 10^7 \text{ E.S.U.}$$



SEA WATER

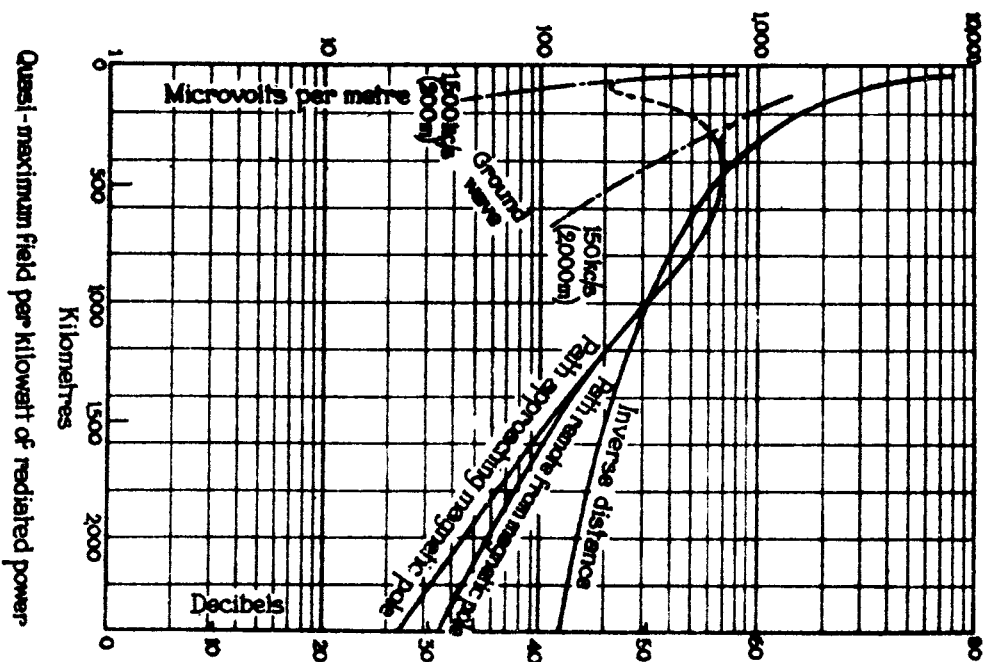
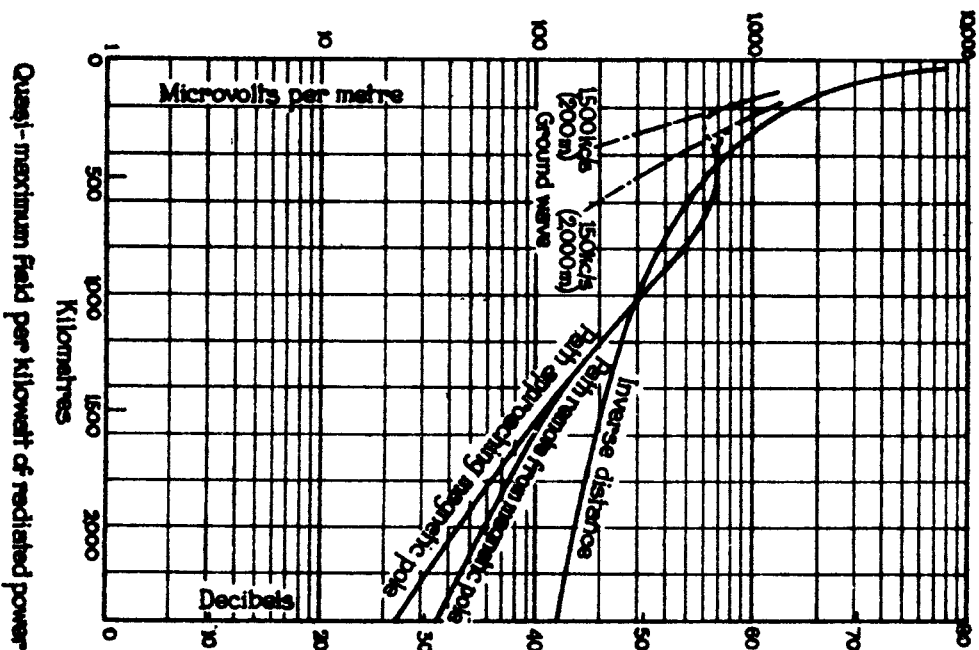
$$\sigma = 3.6 \times 10^{10} \text{ E.S.U.}$$

DAYLIGHT PROPAGATION (GROUND WAVE)

FIG. 18
CHAP. XIV

NIGHT PROPAGATION

FIG. 19
CHAPTER



only about 35 per cent. of the quasi-maximum value. It will be seen that in each graph there are three curves. One, the inverse distance curve, gives the field which would be obtained on a flat earth with geometrical attenuation only, and is inserted merely for comparison. If the transmission is substantially east to west or *vice versa*, in other than equatorial latitudes, the great circle path between transmitter and receiver may approach the magnetic pole, and the absorption will be greater than if the path is remote from the pole. The latter condition corresponds to transmission in the north-south or south-north directions. The ground-wave intensity is approximately shown by the chain-dotted curves. Where the intensities of ground and sky waves are of the same order, the conditions are very variable and the curves are shown in broken line.

Propagation of high-frequency waves (1,500 to 30,000 kc/s.)

67. At frequencies above 1,500 kc/s. the ground wave is rapidly attenuated ; long-distance communication is dependent entirely upon the sky wave, and thus upon the frequency, the height of the reflecting layer, and the electron density in it. The angle at which the wave enters the layer is also of primary importance. The most striking features involved are illustrated in fig. 20,

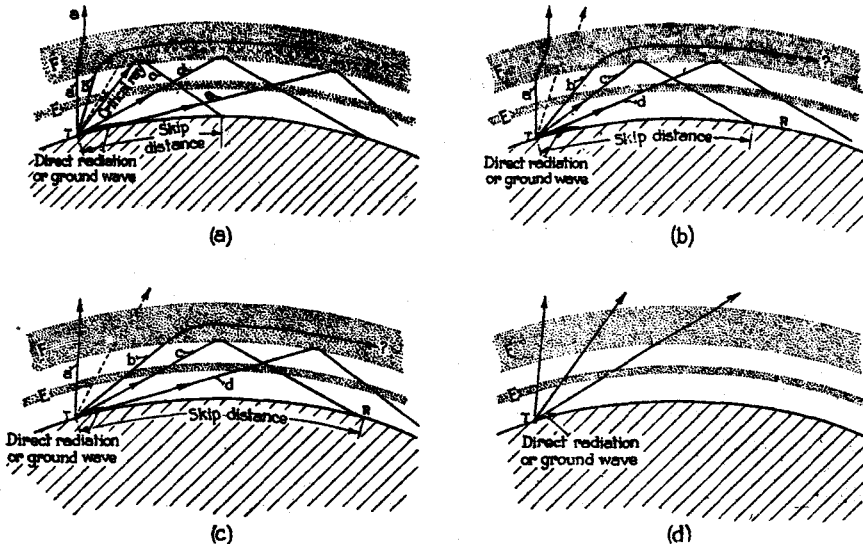


FIG. 20, CHAP. XIV.—Propagation of high frequencies.

which shows the paths of rays of different frequencies emitted at various angles to the vertical. These rays are shown as penetrating the E and reaching the F region. In fig. 20a, the frequency is comparatively low, i.e., in the neighbourhood of 1,500 kc/s. The ray a which is incident almost perpendicularly upon the layer, is subjected to little bending, and passing the region of maximum ionization, escapes past the upper boundary of the layer. Ray b reaches the layer at such an angle that it is just bent parallel to the earth when it reaches the region of maximum density. After travelling parallel to the earth for some distance, it may reach a portion of the layer of greater density at the same height. The ray is then bent downwards and will return to earth. The ray c which leaves the transmitter at a still smaller angle to the ground does not penetrate so far as the region of maximum density but is bent downward much earlier, returning to earth comparatively near to the transmitter. No rays between a and b will return to earth, and in the region between the transmitter and the point at which c returns, only the ground wave will be effective in giving any signal. The ground wave, however, is heavily attenuated, and as a result there is a considerable region in which no signal is received at all ; the distance thus "skipped over" by the sky wave is called the skip distance.

CHAPTER XIV.—PARAS. 68-71

68. In fig. 20b, the frequency is assumed to be somewhat higher, and the refracting property of the layer at this frequency is less than in the previous instance. Ray C, which just reaches the layer and is immediately bent downwards, reaches the ground at a greater distance from the transmitter, i.e., the skip distance is greater. In fig. 20c, a further increase of frequency results in the passage through the layer of almost all the high-angle radiation and ground absorption of almost all the low-angle radiation. The only energy reaching the distant receiver is that reaching the layer at an angle only just above the horizontal. Finally in fig. 20d, the frequency has been increased to such an extent that the F layer is practically non-refracting. No energy is returned to earth by the layer and no long-distance communication is possible. The highest frequency suitable for daylight long-range communication is about 25 Mc/s.

69. It is now seen that for any given frequency there must be a certain critical angle; radiation at an angle greater than the critical, passes through the layer and is lost. The critical angle becomes smaller as the frequency increases, because the refracting property is inversely proportional to the frequency. The effect of the electron density is the reverse; the higher the density, the smaller is the critical angle. The skip distance depends upon the critical angle and upon the height of the layer; it increases as the frequency is raised.

Skip distance/frequency curves

70. It is impossible to state any simple relation between frequency and skip distance. As, however, this question is of first importance in H/F signalling, continuous endeavours have been made to correlate theory and the observed facts, and to produce graphical information which will be of service to those responsible for frequency allocation for different forms of communication. It must be emphasized that such graphs cannot predict with complete certainty the frequency which will give reliable communication between any two points, under any specified conditions of time, date and latitude. Besides the ionospheric variations of a more or less predictable nature already referred to, there are day-to-day changes which appear to be very erratic. Thus a graph drawn to illustrate skip distance/frequency relations over a period may give a rough average picture of these relations for the periods in question, but there will inevitably be certain days when the actual relation differs considerably from those indicated by the graph. Subject to these limitations, the graphs shown in figs. 21 to 33, may be used. These cover the total range of latitude from 65° north to 65° south. The actual zone covered by each graph and the seasons for which it is applicable are shown in a "key page" preceding the graphs. Each graph is made up of seven distance curves marked $d = 0$, $d = 200$ etc., up to $d = 1200$. The distance d is in miles. The ordinate of each graph is marked in frequency (Mc/s) and the abscissa represents local time. It will be observed that five graphs are required for all latitudes other than the zone 15° north to 15° south, which is covered by a single graph for all months of the year. The material on which the graphs for latitudes north and south of 55° are based was very scanty, and these are considered to be less reliable than the remainder.

71. If a short-wave transmitter is communicating with a receiver d miles away, there will be one particular frequency at which waves sent out by the transmitter, and reflected by the ionosphere, will just graze the distant receiver. This is the skip frequency for the distance d miles. With reference to fig. 20c, if the transmitter T is sending on a frequency of f Mc/s and R is the point at which the descending skywave first reaches the earth, then f Mc/s is the skip frequency for the distance T R. If the transmitter T uses a frequency much below f Mc/s, the point R should be well within the zone of the downcoming skywave. Conversely, if the transmitter uses a frequency higher than f Mc/s, the downcoming skywave will strike the earth beyond the point R and no signals will be heard at that point. Local time, at any place, is the time given by a clock set to 1200 hours at the instant when the sun crosses the meridian up to that place. The local time, at any place, may be determined by means of a map, bearing in mind that the change of time for each degree of longitude is 4 minutes; an example of this determination is given below.

Example :—What is the local time at Alexandria, Egypt, when it is 1500 hours B.S.T. in London ?
(1500 hours B.S.T. is 1400 hours G.M.T.)

Reference to a map shows that Alexandria is 30° East of Greenwich and therefore the sun crosses the meridian through Alexandria 30×4 minutes or two hours before it crosses the meridian through Greenwich. 1400 hours G.M.T. at London therefore corresponds to 1600 hours local time at Alexandria, i.e., when it is 1500 hours B.S.T. in London the local time in Alexandria is 1500 hours.

72. When using the skip distance graph to find a suitable frequency for communication between two points such as T and R (fig. 20) the local time to be used is strictly the local time at a point midway between T and R, and not the local time at either T or R. In practice it is seldom necessary to make any correction for the difference in time between the ends and the middle of the signalling path, for even in the worst case—with T and R on the 65th parallel of latitude and 1,200 miles apart—the time difference does not exceed 1.36 hours.

73. The graphs have been so planned that it is possible to use them to estimate approximately (for distances up to 1,200 miles) the skip distance for any given frequency, or alternatively the skip frequency for a given distance. It is probable that the graphs will be used mostly as an aid in finding the most suitable frequency for communication between two known points under specified conditions of time and date, but attention must again be drawn to the fact that it is not possible to predict with certainty the most suitable frequency for reliable communication. If the signalling is to take place between two stations whose geographical position is known, the skip frequency for the distance separating the stations under specified conditions of time and place may be estimated approximately from the appropriate graph. The most suitable frequency for signalling will, however, be below the real skip frequency, for when a transmitter T is sending to a distant receiver at R (fig. 20), the receiver must not be just on the edge of the skip zone, but must lie within the zone of the downcoming skywave, as in fig. 20b. At first sight it might appear that any frequency well below the skip frequency would be suitable for the desired communication. This is not necessarily so, for the attenuation increases as the frequency decreases, so that if too low a frequency is employed, communication may fail because of the excessive attenuation.

74. Where short-wave communication must be established along a certain route and there is local information based on practical experience, concerning the behaviour of short waves along that route, or other routes similar to it, every endeavour should be made to check deduction from the graph by information derived from practical experience. When no local information is available, deduction from the graph should be a valuable guidance to the order of frequency likely to be most useful.

75. In addition to their use as an aid in determining the most suitable frequency for signalling under certain specified conditions, the graphs indicate how frequency conditions along a signalling path are likely to vary with the time of day, e.g., if communication between two points 600 miles apart is satisfactory from 0600 hours to 1000 hours on a certain frequency and then begins to fail, the general trend of the $d = 600$ curves, on the appropriate graph, will indicate whether an increase or decrease of frequency is more likely to enable communication to be satisfactorily maintained after 1000 hours.

Numerical examples

76. Examples illustrating the use of the skip distance frequency graphs are given below. In the earlier examples, no mention is made of any correction to allow for variation of ionospheric conditions with the solar cycle; the method of making such correction is explained later.

Example 1

In fig. 20c, let T represent a short-wave transmitter situated in a place on latitude 20° S., and the distance T R be 400 miles. What will be the skip frequency for the distance T R at 0300 hours, 1000 hours and 1950 hours on a day in November? What frequency would you recommend for communication between T and a receiver at R from 0930 to 1100 hours?

CHAPTER XIV.—PARA. 76

(Example 2)

(a) The skip frequency for any given distance d and time is found by reading off the height in Mc/s of the point where the vertical line corresponding to the time cuts the appropriate d curve for the distance in question. The graph to be used for this problem is No. 6, and the vertical line through 0300 hours cuts the $d = 400$ curve at a height of 5.75 Mc/s. This is the skip frequency at 0300 hours; by applying the same method to the other two times, it will be found that at 1000 hours the skip frequency is 13.5 Mc/s, and at 1950 hours the skip frequency is 15.25 Mc/s.

(b) Reading from the $d = 400$ curve, it will be seen that the skip frequency between 0930 and 1100 hours varies between about 13.2 and 13.5 Mc/s. For reliable communication it is essential to use a frequency below the actual skip frequency; for best results the distant receiver at R must be well within the zone of the downcoming waves, and not just on the edge of the skip zone. On the other hand, the frequency chosen must not be too low; otherwise, if the waves passing from T towards R through the ionosphere are being reflected from the F region, they may be so heavily attenuated that they never reach R at all, or they may reach R but be too weak to give readable signals in the receiver.

(c) It is quite impossible to lay down any rule for the relation between skip frequency and best signalling frequency, but, if no local information is available as a check on results deduced from the skip distance graphs, it would be advisable to try as a signalling frequency, a frequency about 20 per cent. below the deduced skip frequency. Applying this to the case now under consideration, where the skip frequency is 13.2 to 13.5 Mc/s, a frequency of about 10.6 Mc/s (13.2 Mc/s less 20 per cent.) would be recommended as likely to give communication between 0930 and 1100 hours.

Note.—If the frequency 20 per cent. below the deduced skip frequency gave communication, but with signals weaker and less reliable than might reasonably be expected, the indication might be either that the distant receiver was too near the edge of the skip zone, or that excessive attenuation was occurring. Matters might then be improved by an increase or a decrease of frequency, and only experiment could decide which of these two courses would be the better to adopt. If the frequency 20 per cent. below the deduced skip frequency gave no communication at all, the probable indication would be that the deduced skip frequency was too high, and that trials could then be made with frequencies 30 or even 40 per cent. below the deduced skip frequency.

Example 2

A short-wave transmitter situated on a latitude 32° N. sends on a frequency of 16 Mc/s (18.75 m.) at noon on a certain day in January. What skip distance would you expect with this frequency? Estimate the probable skip distances for frequencies of (a) 12 Mc/s (25 m.), (b) 9 Mc/s (33.33 m.) and (c) 6 Mc/s (50 m.), time and date as for the frequency of 16 Mc/s.

Graph No. 7 must be used for this problem. If a point is plotted whose X ordinate is 1200 hours and whose Y ordinate is 16 Mc/s, it will be found to lie on the curve marked $d = 800$. This means that the skip distance for the frequency of 16 Mc/s, is 800 miles at noon (1200 hours).

(a) The point whose co-ordinates are 1200 hours and 12 Mc/s lies about midway between the curves $d = 400$ and $d = 600$ from which it follows that the skip distance for 12 Mc/s at noon is about 500 miles.

(b) The point whose co-ordinates are 1200 hours and 9 Mc/s lies between the $d = 200$ and $d = 400$ curves and is nearer the $d = 200$ curve. By marking in the point and estimating its position relative to both curves, it will be seen that the skip distance for 9 Mc/s at noon is of the order of 250 miles.

(c) If the point whose co-ordinates are 1200 hours and 6 Mc/s is plotted, it will be found to lie below the $d = 0$ curve. This means that at noon the frequency of 6 Mc/s is below the critical frequency of the ionosphere, so that all 6 Mc/s waves sent out by the transmitter, including those travelling vertically upwards, are reflected back to earth. At noon, therefore, the skip distance for the frequency of 6 Mc/s is zero.

Example 3

A short-wave transmitter situated on latitude 20° N. sends on a frequency of 5 Mc/s (60 m.) throughout the whole of a 24 hours' period in February. How will skip distance vary with time throughout the 24 hours?

If a ruler is laid across graph No. 3, parallel to the time axis, and at a height of 5 Mc/s, it will be seen that the horizontal line for 5 Mc/s starts on the $d = 0$ curve, cuts twice across the $d = 200$ curve, then cuts across the $d = 0$ curve and remains below this curve thereafter. Translating these results into words, there is no skip distance at midnight, but from midnight to about 0315 hours skip distance increases steadily until it reaches a maximum value of about 220 miles at 0315 hours. It then starts to decrease and by 0645 hours has fallen to zero. From 0645 hours to midnight skip distance remains at zero, i.e., there is no skip effect during these hours.

Example 4

A comparatively low power short-wave transmitter in Palestine is working in January to a receiver situated 700 miles east of it. Communication is opened on a frequency of 4.5 Mc/s at midnight and remains satisfactory until about 0400 hours, when the distant receiver reports that signals are falling in strength and becoming unreadable. To what can this falling-off of signals be attributed, and would an increase or decrease of wavelength be advisable if it be required to continue communication until 0700 hours?

From a reference to a map and the time of the year (January) it is obvious that graph No. 7 must be used for this problem. By sketching in a " d " curve for 700 miles intermediate between the curves $d = 600$ and $d = 800$, it can be seen that the skip frequency for 700 miles is about 6.5 Mc/s at midnight, after which it increases steadily. At 0400 hours it has reached nearly 9 Mc/s and is rising fairly steeply. Since communication is satisfactory up to 0400 hours on the frequency of 4.5 Mc/s, the indication is that the falling-off of signals which starts at about 0400 hours is not the result of using too high a frequency; i.e., the downcoming waves are not coming down to earth beyond the receiver. It may, therefore, be assumed that the falling-off of signals is due to the fact that the frequency of 4.5 Mc/s is being increasingly attenuated in its passage through the ionosphere, so that in order to maintain communication from 0400 to 0700 hours, it would be advisable to change to some higher frequency nearer to the actual skip frequencies of 9 to 12.5 Mc/s.

Example 5

Communication is required between a short-wave station at Bombay (Lat. 19° N.) and another similar station at Calcutta, which is approximately due east of Bombay. Both stations have an aerial power of about 0.5 kilowatts. What frequencies might be tried as likely to give good results from 0900 to 1200 hours, and from 2100 to 2400 hours (Bombay local time), during the months of May, June, and July?

Calcutta is approximately 1,100 miles from Bombay, from which it differs in longitude by about 18° . There is, therefore, a time difference of 72 minutes, and the local times at the mid-point of the signalling path corresponding to the Bombay times specified above will be 0936 to 1236 hours and 2136 to 0036 hours.

(a) By using graph No. 6 and interpolating between the $d = 1,000$ and $d = 1,200$ curves, the skip frequencies for different times within the required first signalling period are found to be approximately as follows:—

0936 hours—20.5 Mc/s.
1036 hours—21.0 Mc/s.
1136 hours—20.75 Mc/s.
1236 hours—20.0 Mc/s.

CHAPTER XIV.—PARA. 77

It is obvious that some frequency below 20 Mc/s is indicated as the signalling frequency, and, if no local information were available as a check on the figures above, a frequency of 16 Mc/s (20 Mc/s less 20 per cent.) might be tried for communication during the period in question (but also see Note to Example 1).

(b) For the second signalling period the skip frequencies deduced from the graph are as follows:—

2136 hours—22·5 Mc/s.
2236 hours—20·0 Mc/s.
2336 hours—17·0 Mc/s.
0036 hours—13·75 Mc/s.

(c) The rapidly falling skip frequency makes it unlikely that, with the small power available, one frequency could be used successfully for the whole period under consideration. Communication might probably be started on a frequency of about 16 Mc/s (20 Mc/s less 20 per cent.), and if this were found to fail, very possibly between 2230 and 2330 hours, a change could then be made to a frequency in the region of 11 Mc/s, i.e., a frequency about 20 per cent. below 13·75 Mc/s (also see Note to Example 1).

Corrections for sun-spot cycle

77. It was stated in an earlier paragraph that ionization density changes progressively throughout the period of a solar cycle. As a solar cycle proceeds from a condition of minimum sun-spot activity to a condition of maximum sun-spot activity, ionization density increases, and therefore there is a progressive decrease in the value of skip distance for any given frequency. The skip distance for a given frequency rises to its highest value when the sun-spot activity is a minimum and falls to its lowest value when the sun-spot activity is a maximum. Conversely, the skip frequency for any given distance has its highest value when sun-spot activity is a maximum, and its lowest value when sun-spot activity is a minimum.

Percentage Additions to Frequency to Allow for Solar-Cycle Effects

Year	Latitude					
	10°	20°	30°	40°	50°	60°
1930	3	6	9	12	15	18
1931	2	4	6	8	10	12
1932	1	2	3	4	5	6
1933	0	0	0	0	0	0
1934	0	0	0	0	0	0
1935	1	2	3	4	5	6
1936	2	4	6	8	10	12
1937	3	6	9	12	15	18
1938	4	8	12	16	20	24
1939	5	10	15	20	25	30
1940	4	8	12	16	20	25
1941	3	6	9	12	15	18
1942	2	4	6	8	10	12
1943	1	2	3	4	5	6
1944	0	0	0	0	0	0
1945	0	0	0	0	0	0
1946	1	2	3	4	5	6
1947	2	4	6	8	10	12
1948	3	6	9	12	15	18
1949	4	8	12	16	20	24

The skip distance graphs have been plotted for a time of minimum sun-spot activity, which last occurred about 1933/1934, but by using the above table it is possible to make approximate corrections for the progressive ionospheric changes which have occurred since that time. Notice that the percentage additions to frequency which must be made to allow for the solar cycle vary with latitude; the higher the latitude, the greater is the change of ionospheric conditions resulting from changing solar conditions. Two examples illustrating the use of the correction table are given below.

Example 6

In 1933/34 signalling along a path 1,000 miles long and lying between latitudes 25° and 35° N. was carried out between 1000 and 1300 hours every day in February by using a frequency of 13 Mc/s. What frequency should be employed for similar work in 1937?

On reference to the table, it will be seen that for latitudes 30° a 9 per cent. addition to the 1933/34 frequency should be made to allow for the difference between ionospheric conditions in 1937 and 1933/34, i.e., the 1937 frequency should be $13 \text{ Mc/s} + 1.17 \text{ Mc/s} = 14.17 \text{ Mc/s}$.

Example 7

Correct the answers to (a), (b) and (c) of Example 2, the corrected answers to give skip distances for the year 1937 for the frequencies of 12, 9 and 6 Mc/s.

(a) A frequency of 12 Mc/s in 1937 is equivalent to a frequency of f Mc/s in 1933/34, where

$$f + .09f = 12$$

$$\text{i.e., } f = \frac{12}{1.09} = 11 \text{ Mc/s.}$$

From graph No. 7 the skip distance at noon for this frequency is approximately 400 miles.

Applying the reasoning above to cases (b) and (c) of Example 2, the 1933/34 frequencies corresponding to 1937 frequencies of 9 and 6 Mc/s are 8.25 Mc/s and 5.5 Mc/s, and these give skip distances at noon of just under 200 miles and zero respectively. Hence, for 1937, the skip distances at noon for the three frequencies are :—

- (a) 12 Mc/s—400 miles (approx.).
- (b) 9 Mc/s—200 miles (approx.).
- (c) 6 Mc/s—zero.

SKIP-DISTANCE/FREQUENCY GRAPHS

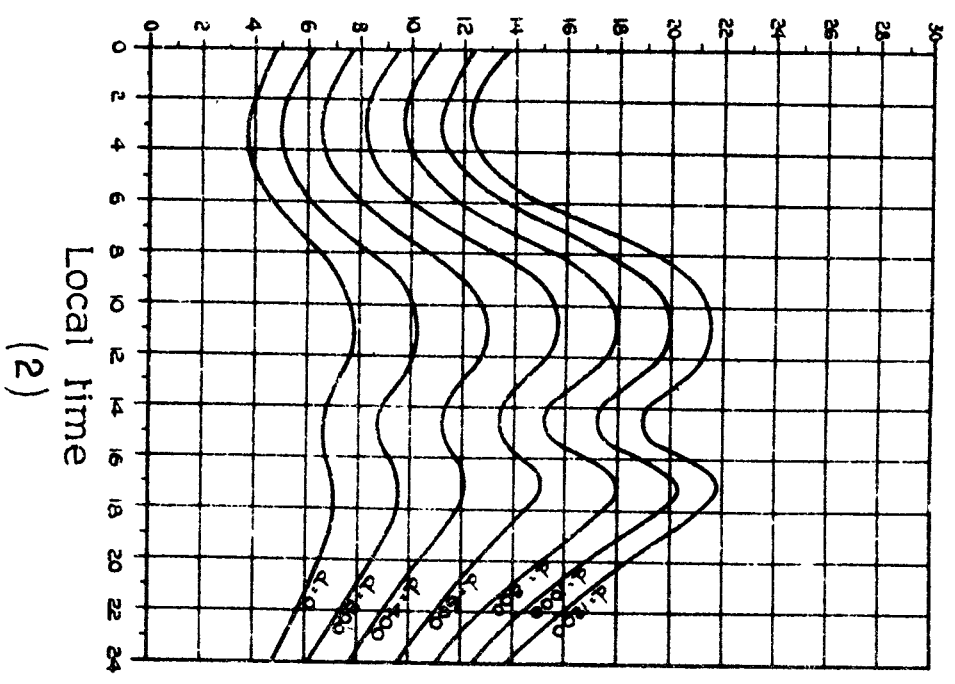
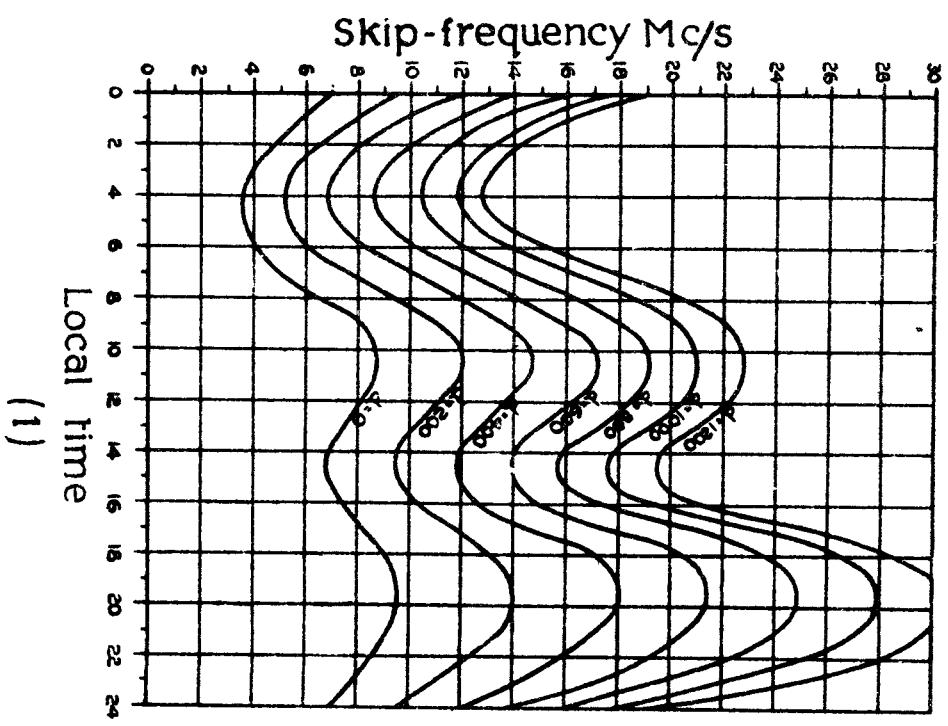
(i) The graphs are numbered consecutively from 1 to 26, each diagram, figs. 21 to 33 inclusive, containing two graphs.

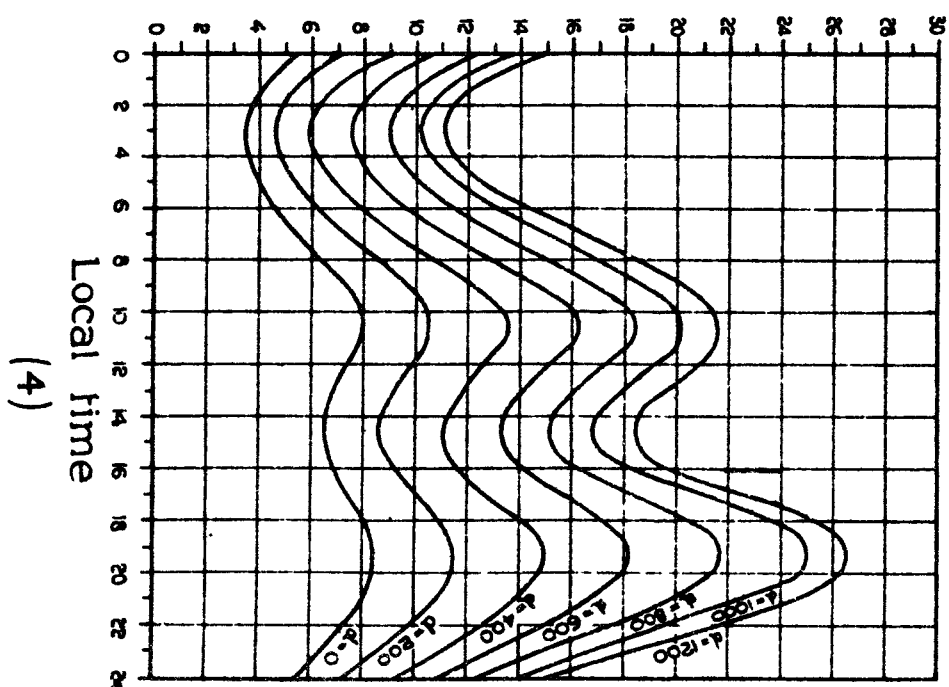
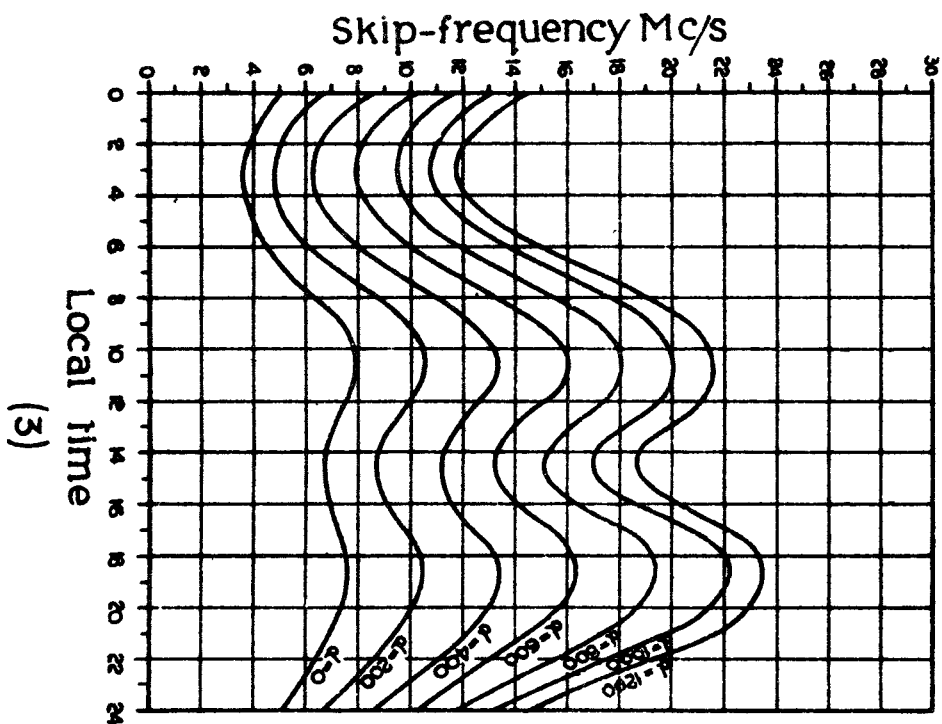
(ii) The appropriate graph number for any latitude and month is given in the following table :—

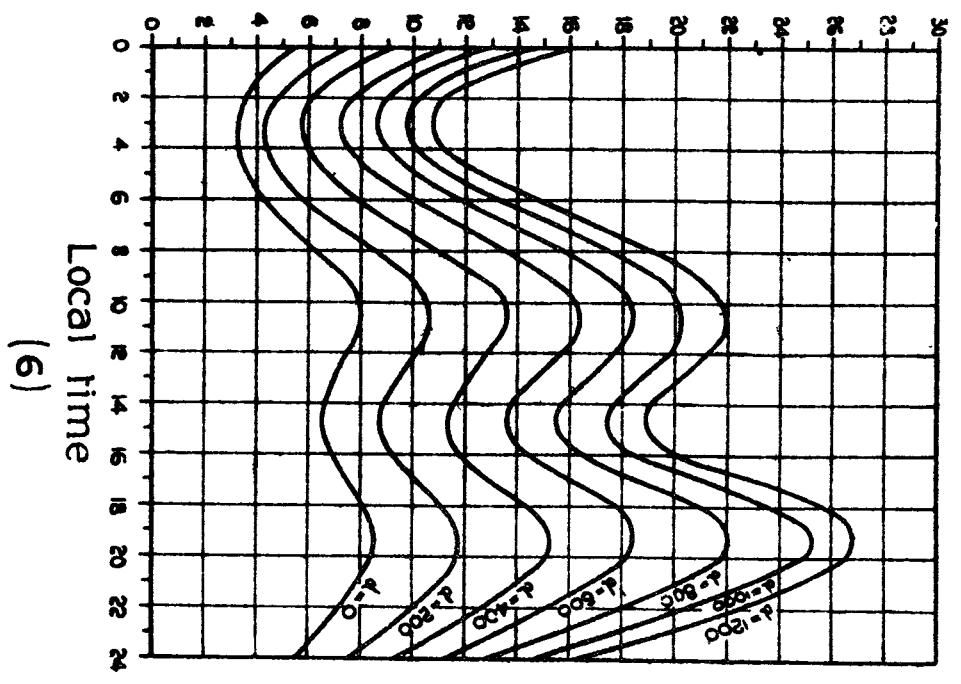
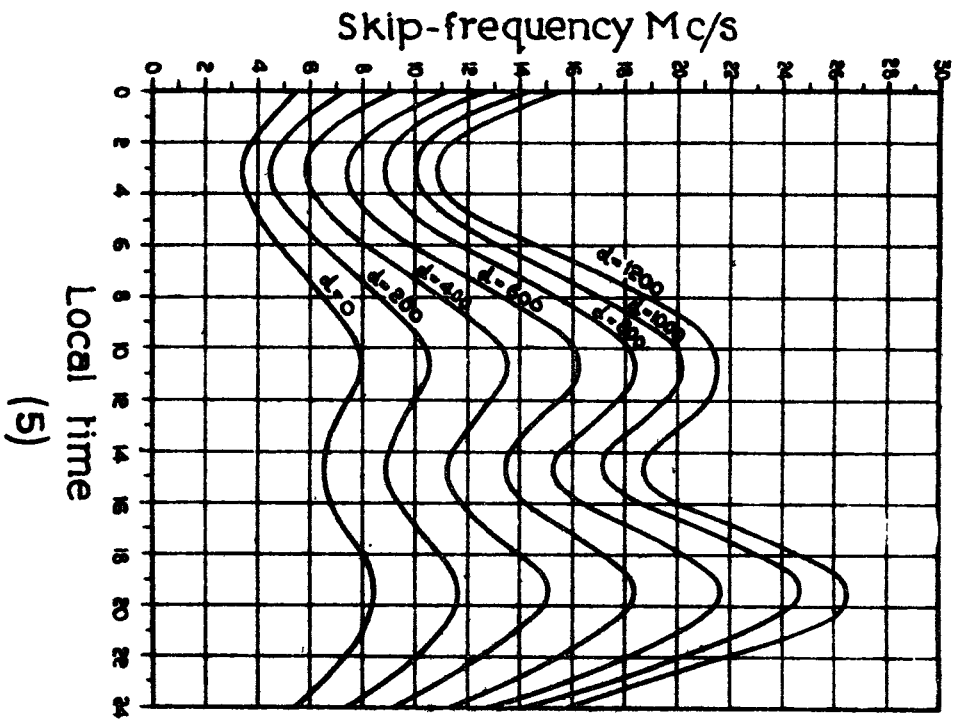
Hemisphere.		Lat. 15° N. to 15° S.	Lat. 15° to 25° N. or S.	Lat. 25° to 35° N. or S.	Lat. 35° to 45° N. or S.	Lat. 45° to 55° N. or S.	Lat. 55° to 65° N. or S.
Northern	Southern						
Jan., Nov., Dec.	May, June, July ..	1	2	7	12	17	22
Feb., Oct. ..	April, Aug. ..	1	3	8	13	18	23
March, Sept. ..	March, Sept. ..	1	4	9	14	19	24
April, Aug. ..	Feb., Oct. ..	1	5	10	15	20	25
May, June, July	Jan., Nov., Dec. . .	1	6	11	16	21	26

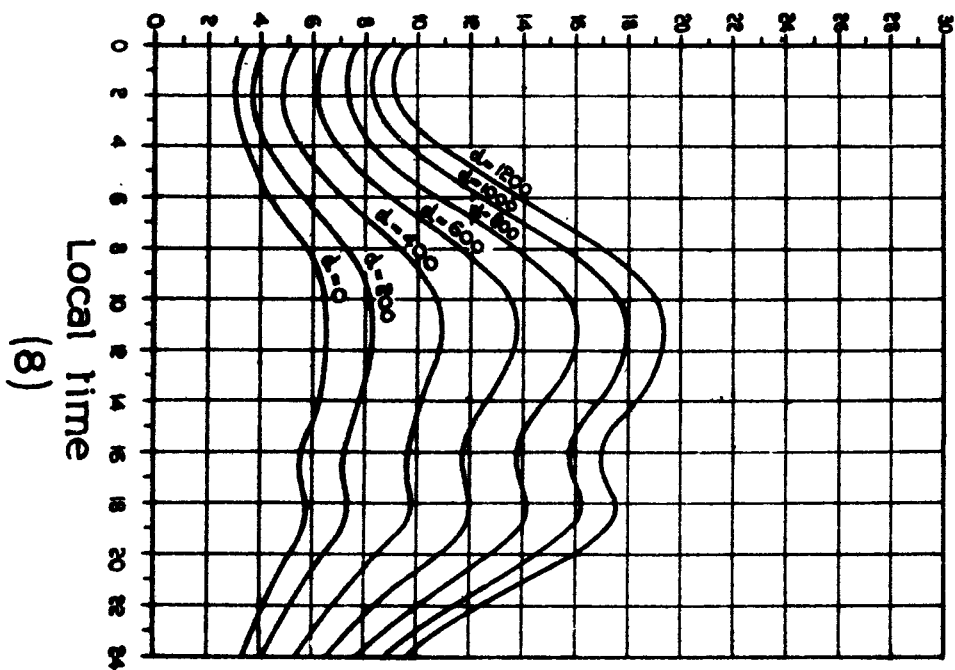
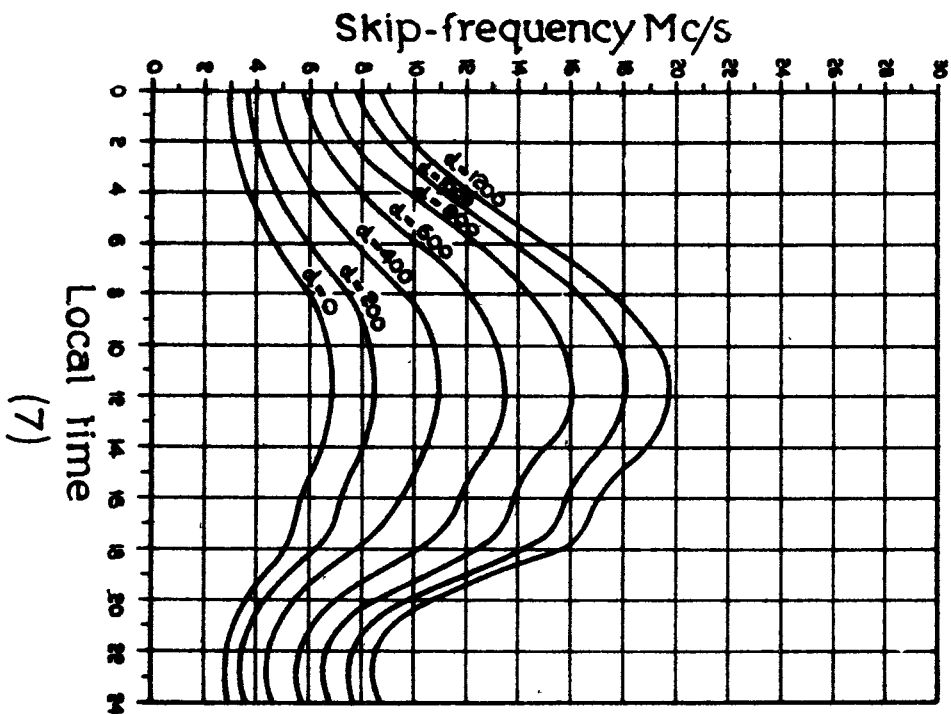
(iii) The graphs are calculated for the minimum period of a sun-spot cycle ; corrections must be applied for other periods.

(iv) The range " *d* " is in miles.









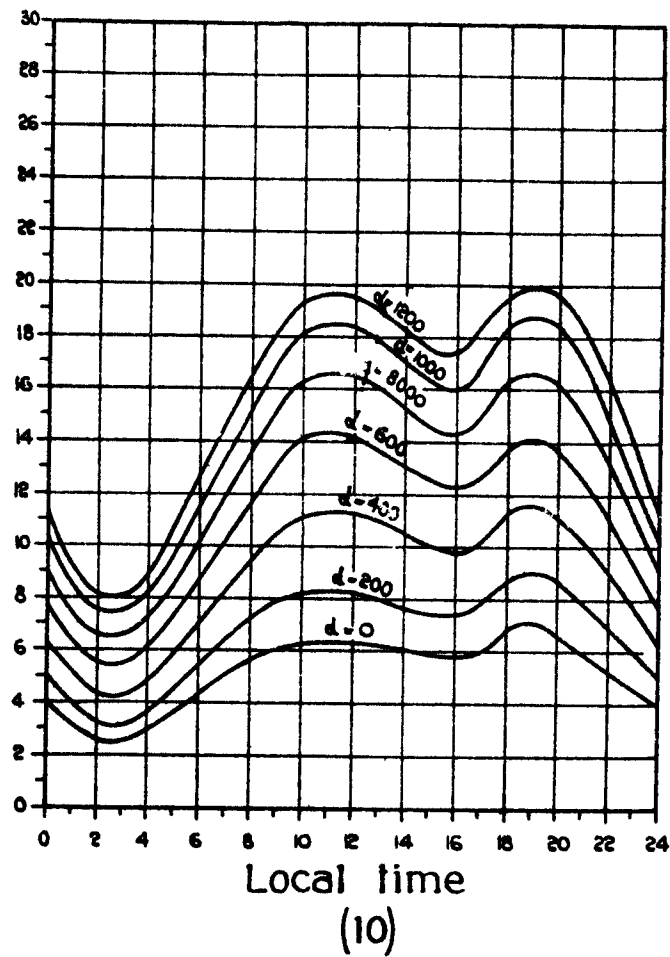
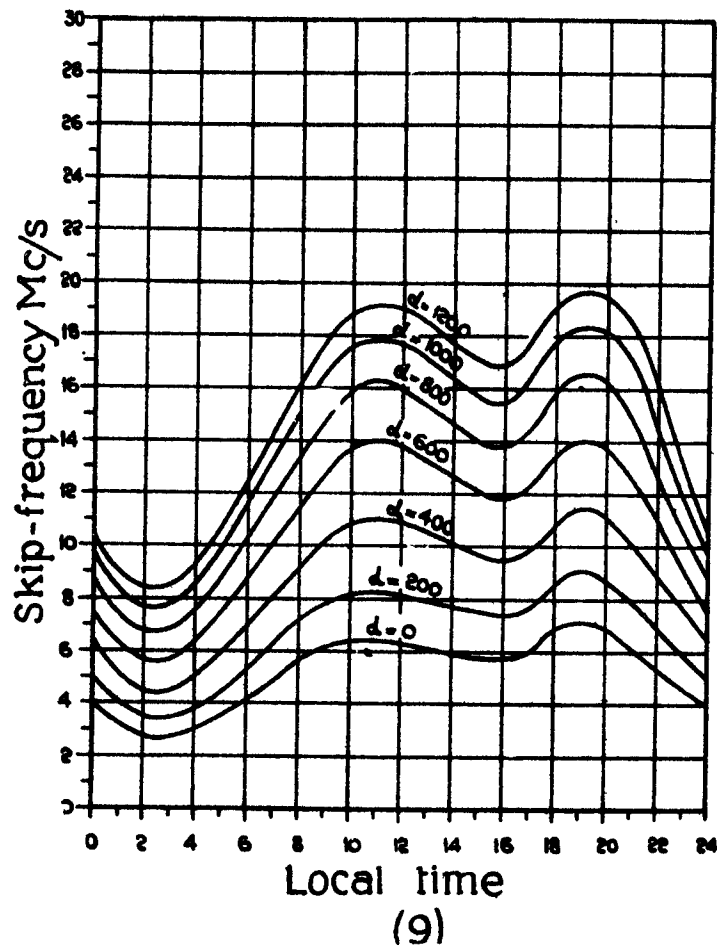


FIG. 25
CHAP. XIV

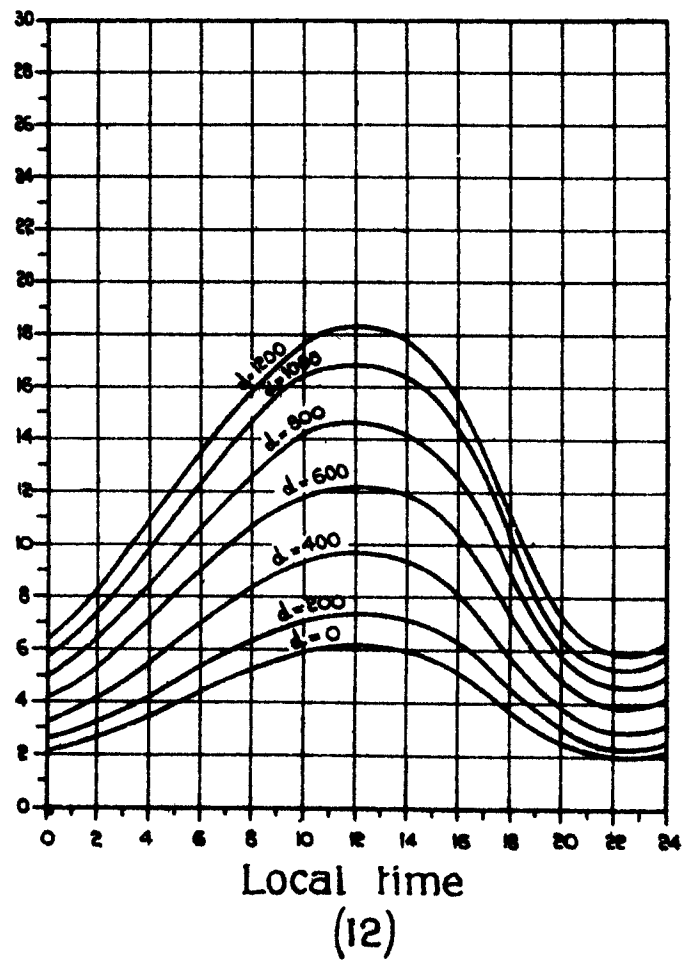
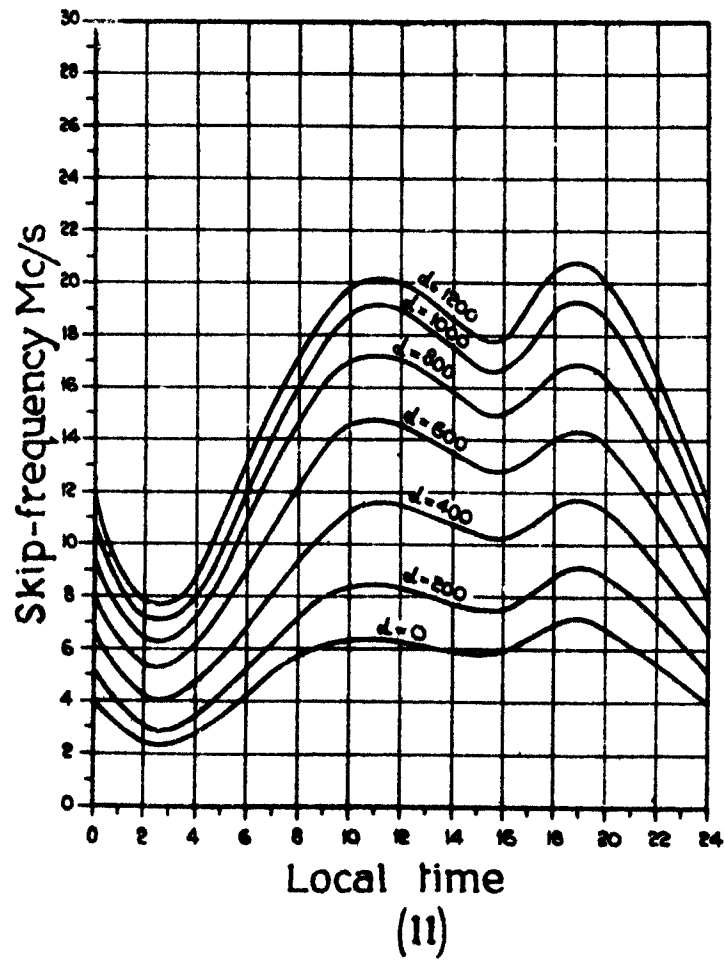
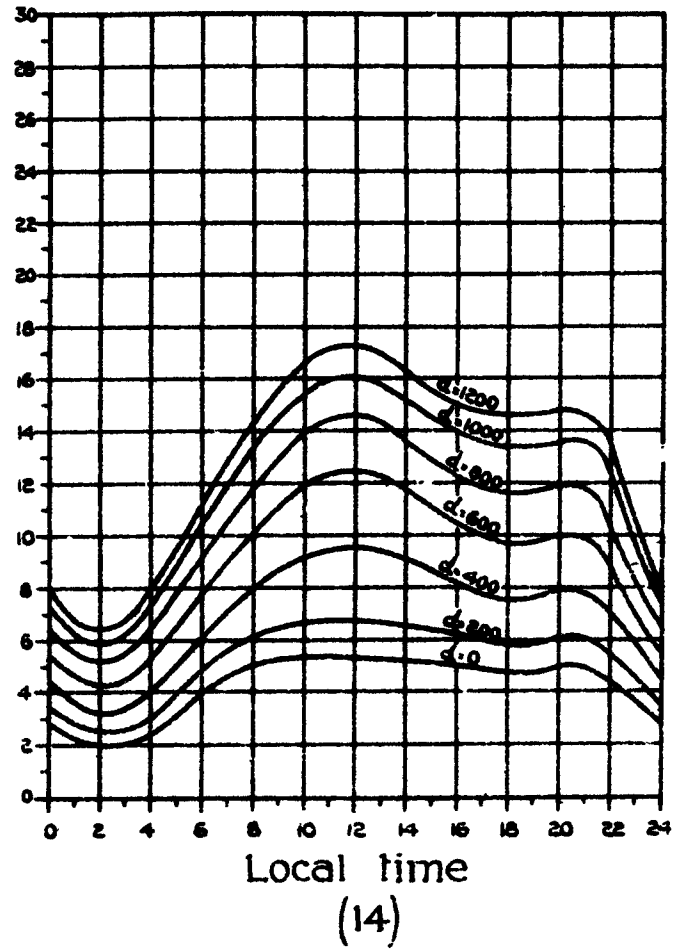
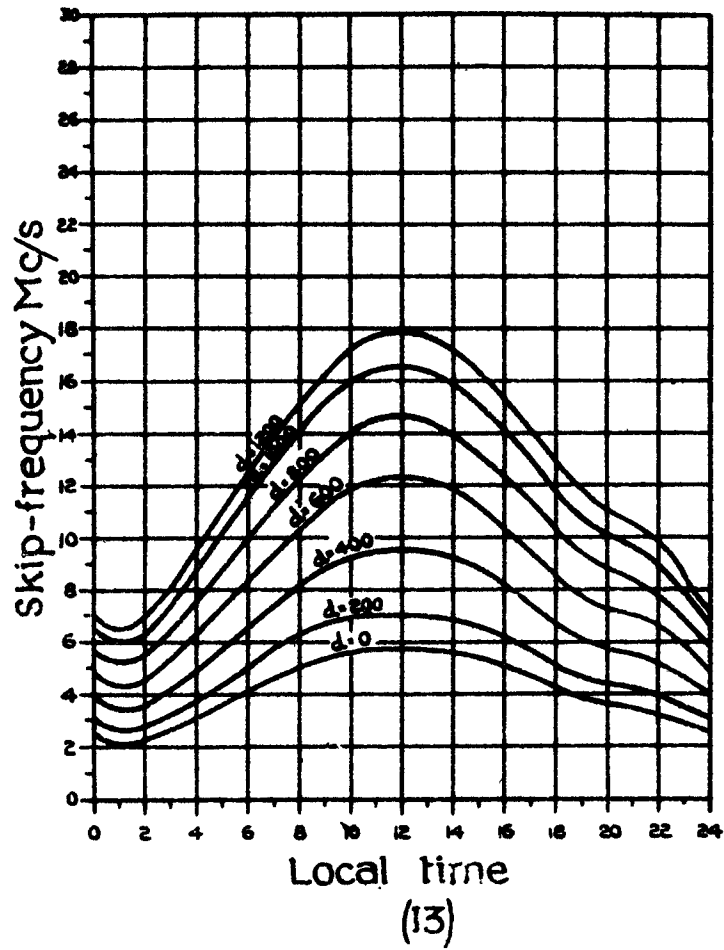


FIG. 26
CHAP. XIV



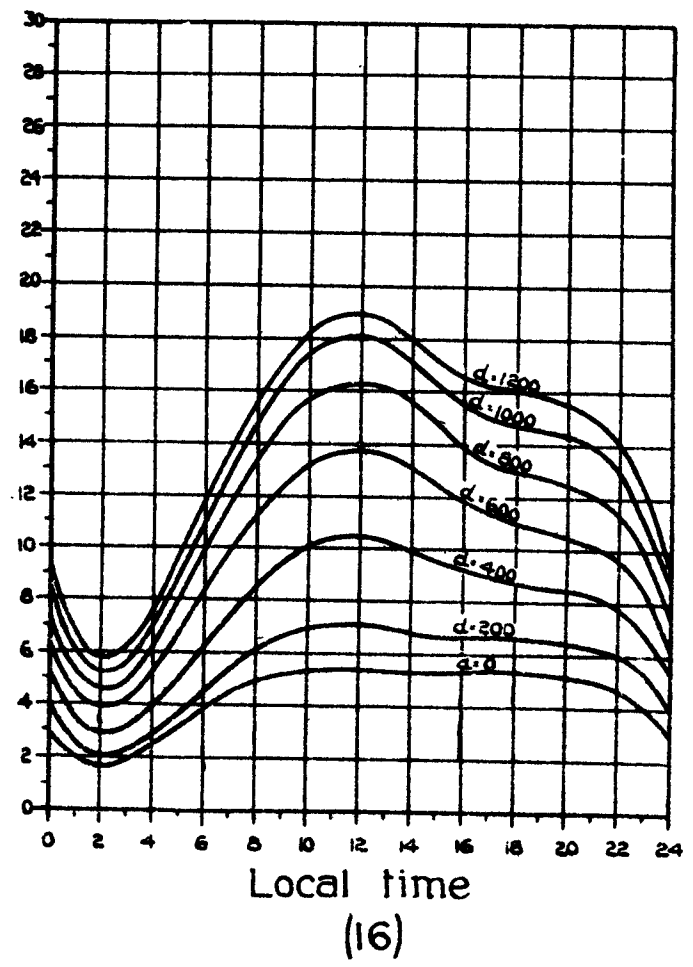
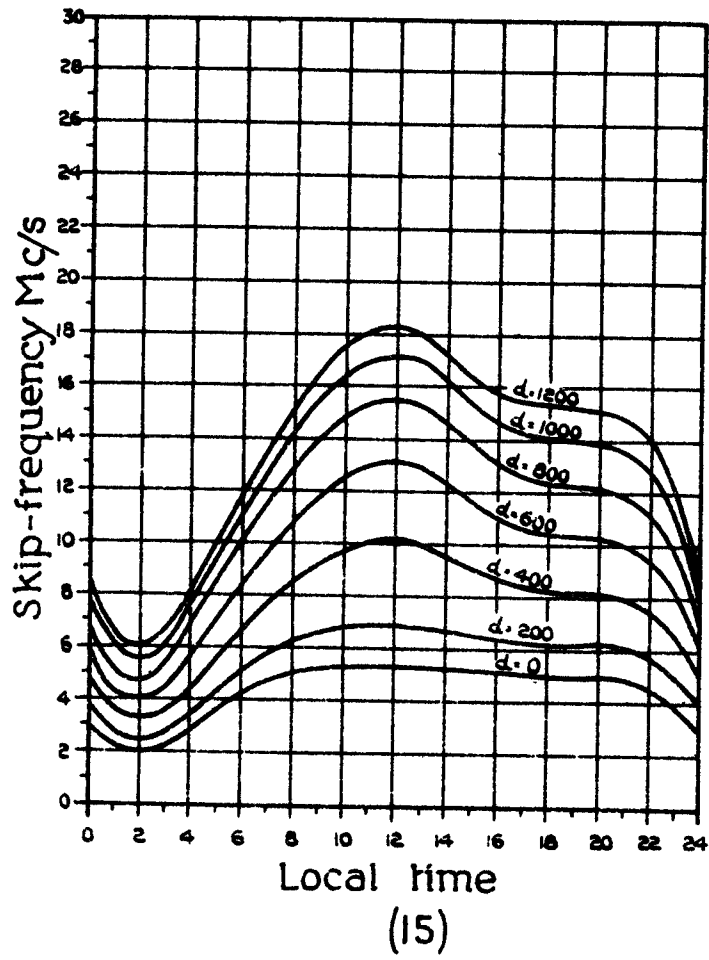


FIG. 28
CHAP. XIV

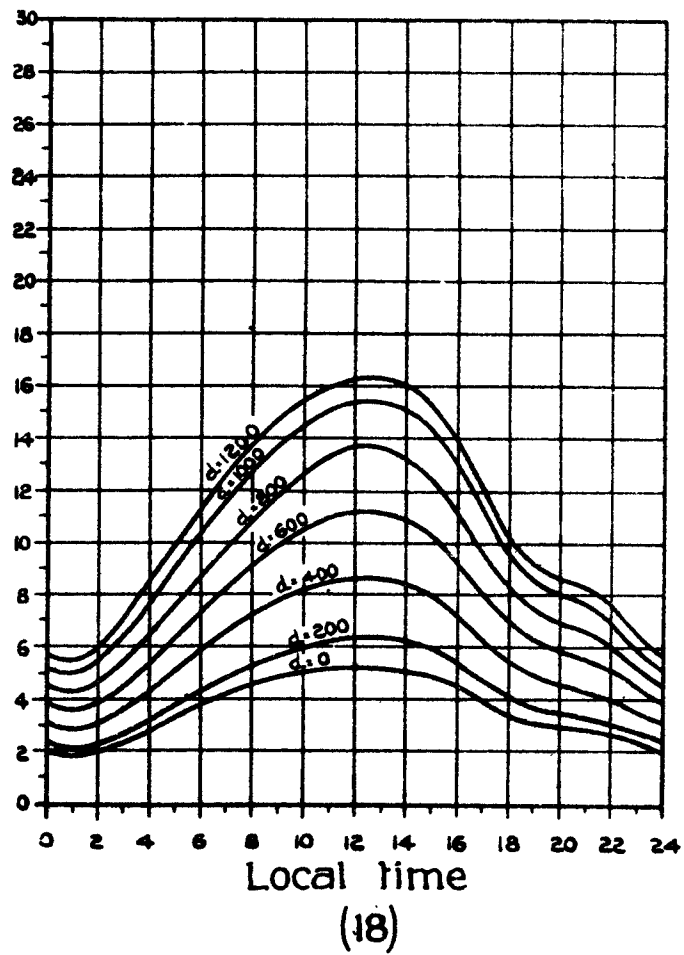
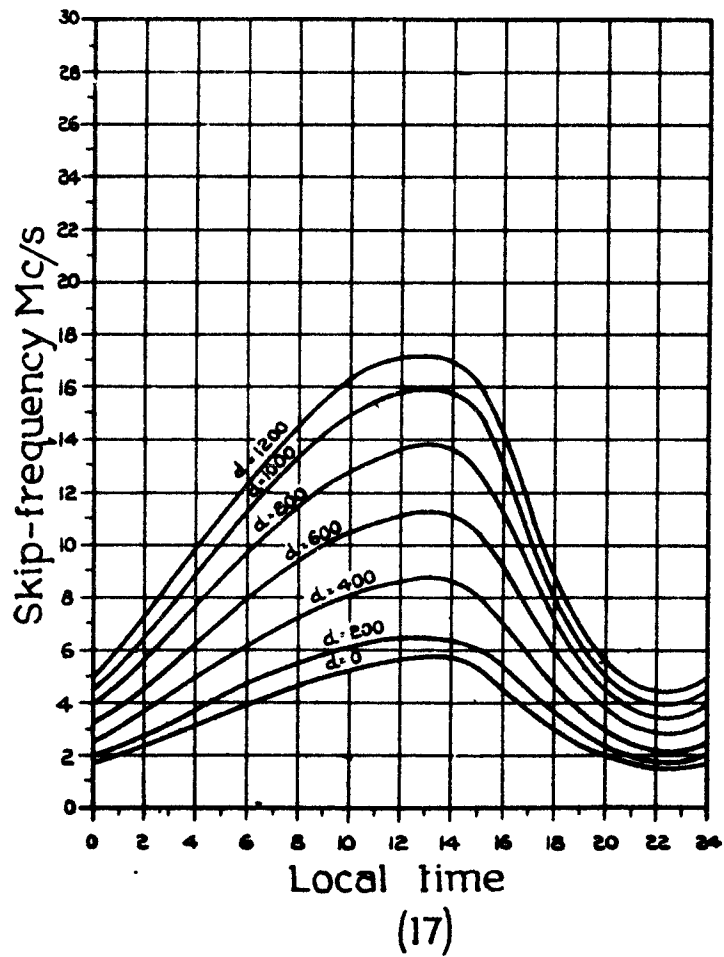


FIG. 29
CHAP. XIV

FIG. 30
CHAP XIV

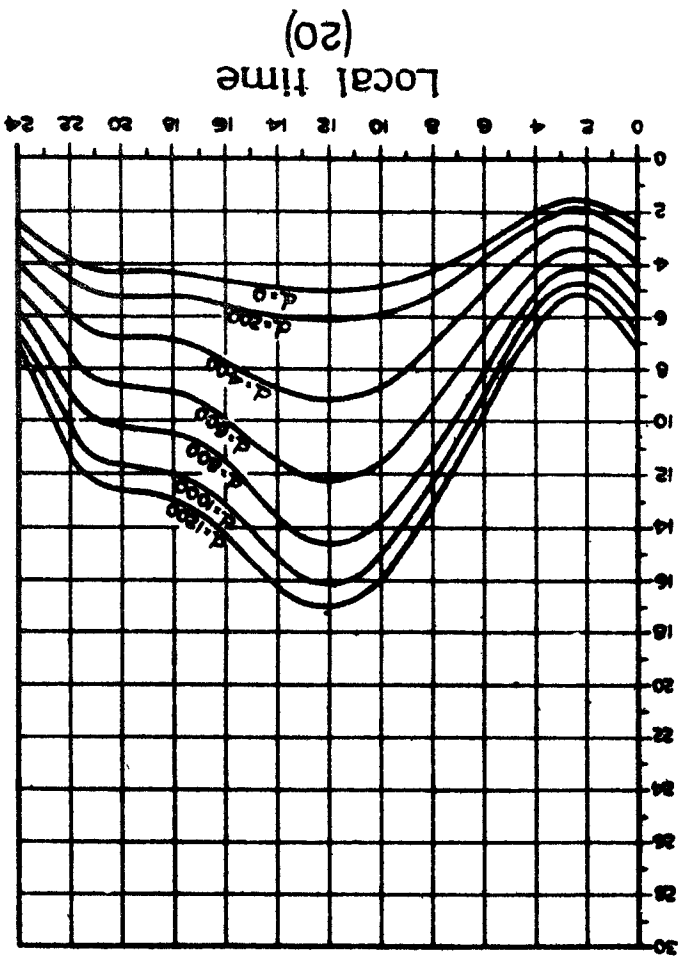
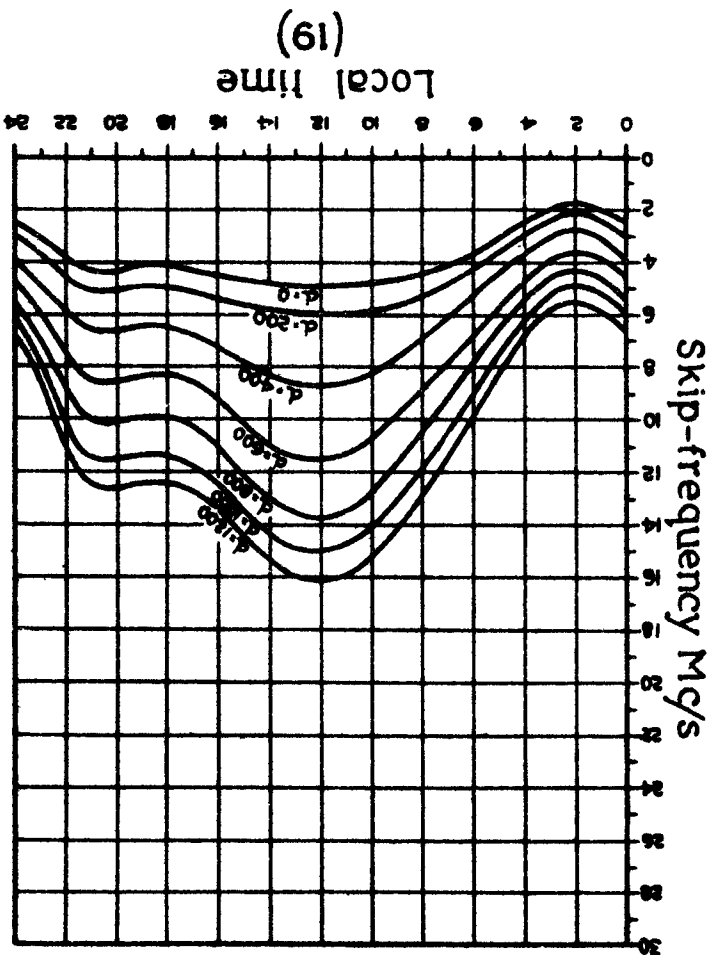
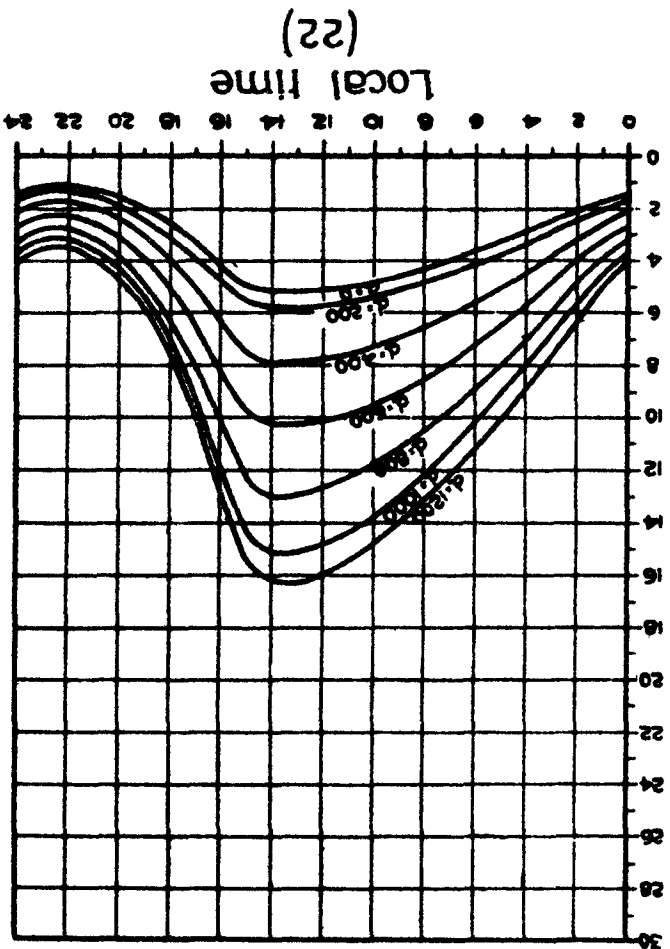
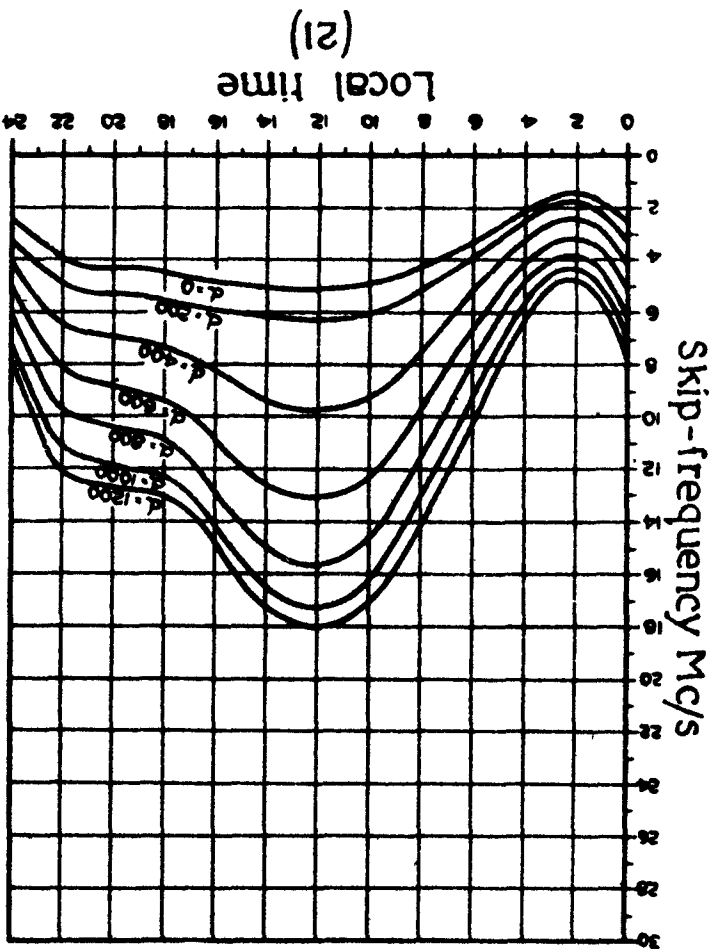
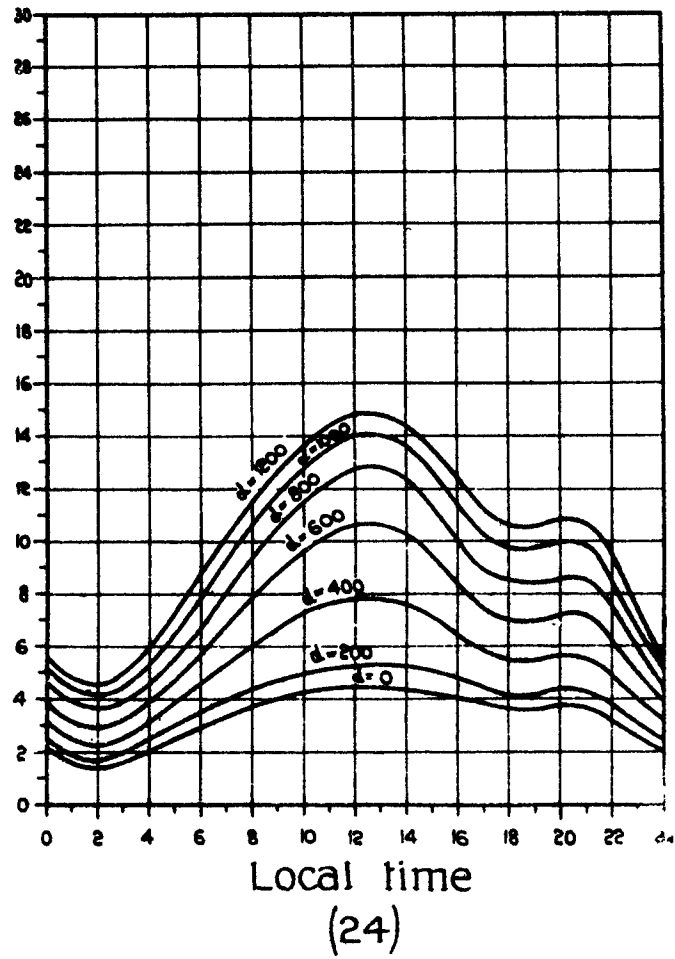
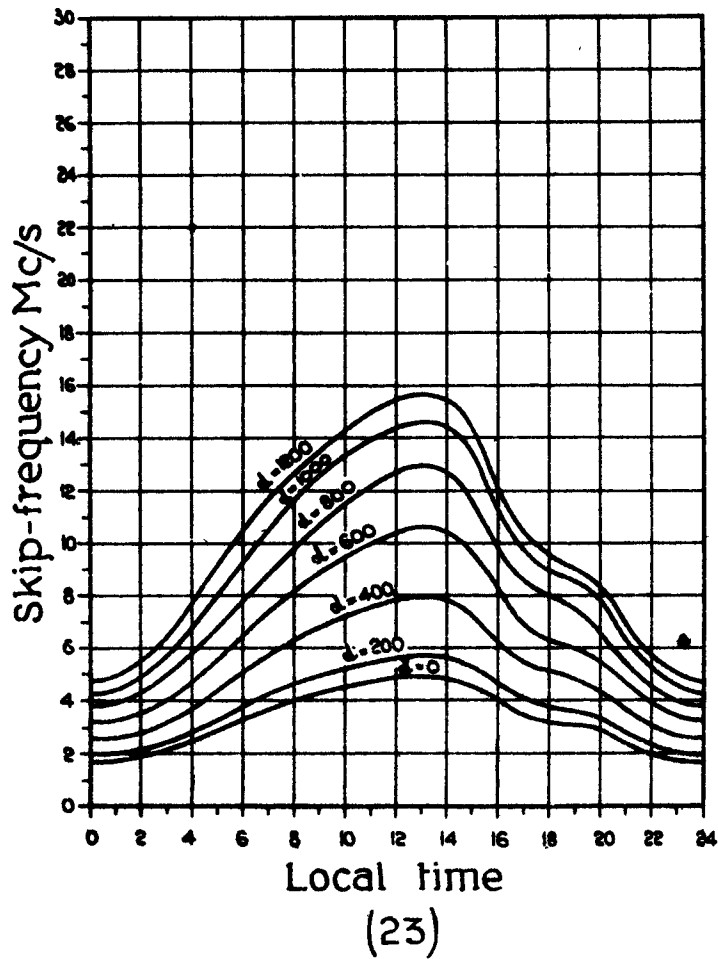


FIG. 31
CHAP. XIV





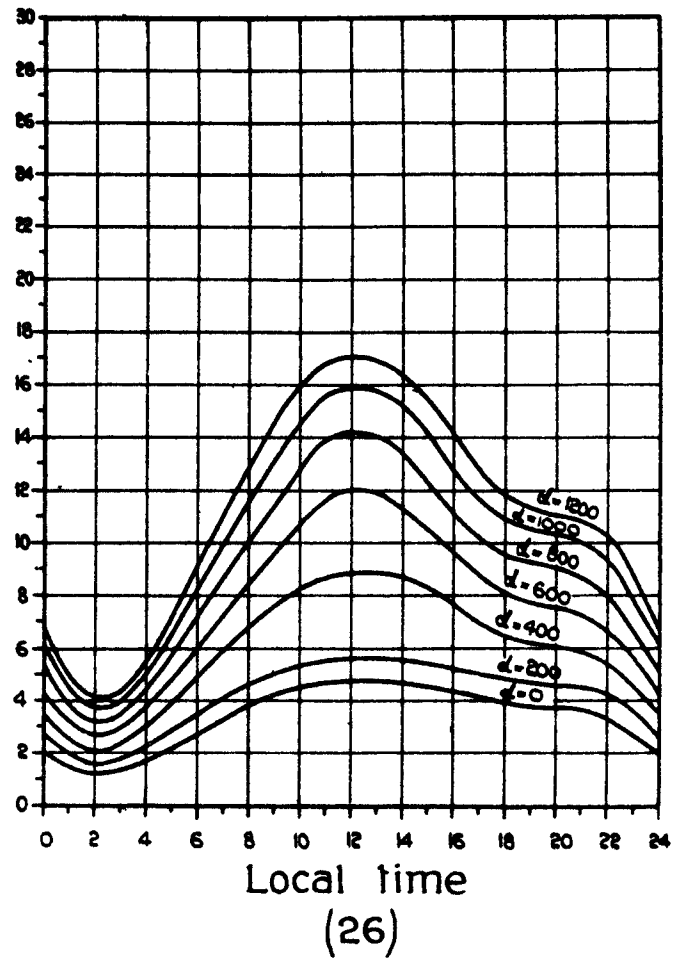
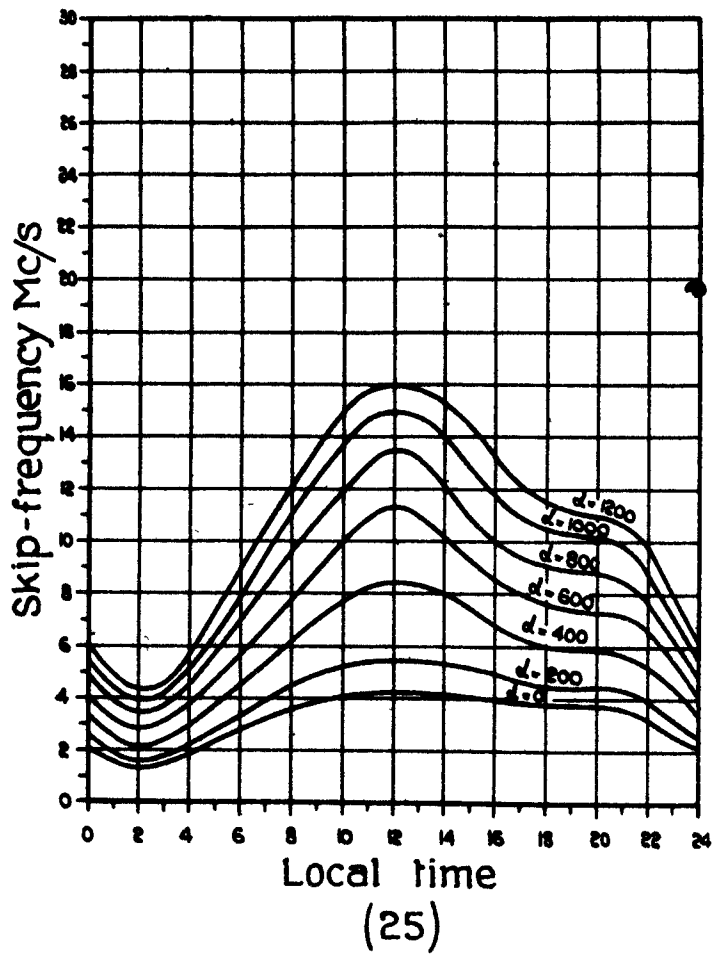


FIG. 33
CHAP. XIV

CHAPTER XV.—AERIAL ARRAYS AND TRANSMISSION LINES

AERIALS AND AERIAL ARRAYS

Introductory

1. From the earliest days of radio communication, the advantages of directional transmission and reception, particularly for the purpose of point-to-point communication, have been fully appreciated. The earliest attempts in this direction were made at comparatively low frequencies (of the order of 15 to 30 kc/s) using L aeriels having a great horizontal length—some ten to twenty times the height. Such aeriels were very expensive in first cost and maintenance. With the development of high frequency communication, the employment of highly directional aerial systems proceeded rapidly. For any energy-radiating system to possess directional properties, its dimensions must be at least comparable with the wavelength of radiation in the particular medium. For example, suppose it is desired to radiate a beam of sound waves at a frequency of 500 cycles per second by means of a horn. As the speed of sound in air is about

1,120 feet per second, the wavelength is $\frac{1,120}{f}$ or 2.24 feet, and the mouth of the horn, if square,

should be at least 2 feet by 2 feet, and if circular or elliptical it should have an area of at least 4 square feet. In the same way, directional electro-magnetic radiation requires that the aerial system shall have dimensions of at least the same order as the wavelength, and this requirement is obviously more easily met at high frequencies (short wavelengths) than at the low frequencies formerly employed.

2. A combination of radiating members designed for the purpose of directional transmission or reception is called an aerial array. The object of an aerial array is to produce some particular distribution of field strength in space, according to the nature of the service, the distance of the receiving station and other factors. The spatial distribution of field strength may be shown by horizontal and vertical polar diagrams as in the case of single aeriels. Aerial arrays are for the most part employed in long distance point-to-point communication, and for this service the horizontal polar diagrams should be long and narrow. Since the propagation is dependent upon reflection from the ionosphere the vertical polar diagram should be such that most energy is radiated at a low angle to the horizontal, usually between ten and fifteen degrees. It is found that the apparent direction from which the strongest radiation arrives at a receiver depends partly upon the state of the ionosphere, and may vary through several degrees from hour to hour or from day to day. It is therefore not desirable to aim at an extremely directive polar diagram. At the receiver, a fairly sharp vertical diagram is an advantage, provided that the optimum angle can be decided, because under these conditions less trouble is caused by echo phenomena. In general, however, the optimum angle also varies with the time of day, season, etc. Certain special types of communication, e.g. ground to air and *vice versa*, may require types of array very different from those used for long distance point-to-point communication.

Reciprocal properties

3. The properties of any aerial, when used for reception, are in most respects similar to the corresponding properties of the same aerial when used for transmission. In particular, the directional characteristic is practically unchanged. The current distribution and effective impedance are not quite the same because the current is due to a field spatially distributed over the whole aerial (not necessarily in a uniform manner) instead of an E.M.F. applied between two feeding points, and the impedance changes slightly owing to an indirect effect of the current distribution. The fact that the directional characteristic is substantially the same enables the merits of a given aerial or array for transmitting purposes to be deduced from its behaviour as a receiver and *vice versa*.

CHAPTER XV.—PARA. 4

Current distribution

4 The current and voltage distributions along an aerial wire were dealt with briefly in Chapter VII. It is now necessary to discuss the current distribution somewhat more fully. Consider an aerial suspended above the earth in any manner whatever, e.g. as shown in fig. 1a, and its lower end to be connected to one terminal of a high frequency generator, the other terminal of which is earthed. In order to measure the amplitude of the current at different points of the wire, we may use a thermo-ammeter, inductively coupled to the aerial by means of a loop of wire. This device is in fact in constant use for the adjustment of aerial arrays, and suitable dimensions are given later. If arrangements are made to draw this loop along the aerial we

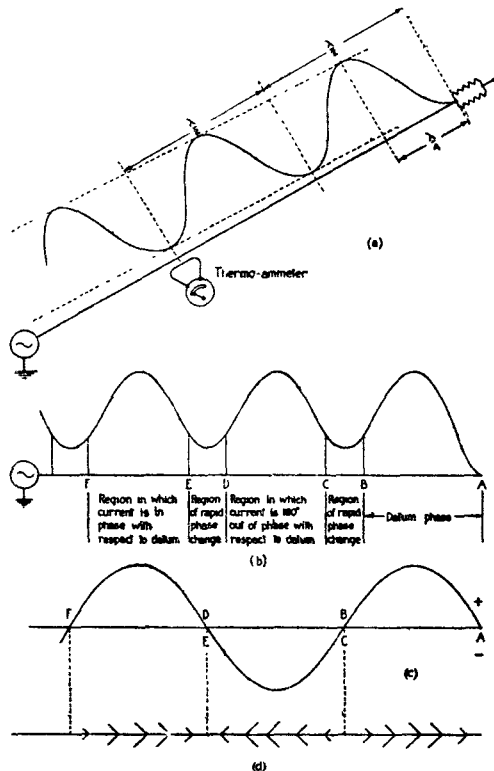


FIG. 1, CHAP. XV.—Current distribution along wire.

may obtain an indication of the R.M.S. current at different points. At the end remote from the generator, the current in the wire will be zero. As the loop is moved towards the generator the current increases, and will be found to pass through a maximum at a distance exactly $\frac{\lambda}{4}$ from the open end. It then decreases, and passes through a minimum value at a distance of $\frac{\lambda}{2}$ from the end. It will then start to increase again and will pass through another maximum when the ammeter is $\frac{3}{4}\lambda$ from the end, this maximum being slightly greater in amplitude than the previous one. If at each point in the wire a perpendicular is drawn, and its length is proportional to the current at that point the ends of these lines will lie upon a curve as in fig 1a. This curve then gives the current distribution, so far as its magnitude is concerned, but it will

give no information as to the relative phase of the current at different points in the wire. If, however, steps were taken to measure this it would be found that starting from the far end, the current at all points in a length A B, i.e. over a distance of nearly $\frac{\lambda}{2}$, is very nearly of the same phase. Over a short length B C, fig. 1b, in the vicinity of the current minimum, the phase changes very rapidly, and in passing from B to C a total change of 180° takes place. In the distance C D, the current is syn-phased at all points and is therefore 180° out of phase with the current in A B. This process of phase reversal again occurs in the length D E, so that the current in E F is in phase with that in A B, but opposite in phase to that in C D.

5. In Chapter VII it is assumed that the lengths B C, D E, etc., are so small that they may be represented by geometrical points, and also that the current at these points, instead of being a minimum, falls to zero. This assumption is often made in theoretical work, because under these conditions both the magnitude and the phase of the current can be shown by a sine curve, fig. 1c; at points lying above the wire the current is of the same phase throughout, and at points lying below the wire the current is 180° out of phase with the points lying above it. Alternatively, we may say that if at any instant the current is flowing from B to A, the current in B C is flowing towards B, in the length C D from C to D and so on. We may therefore show the distribution of current along the wire by drawing a series of arrows of varying sizes as in fig. 1d. This method is of great assistance when considering the distribution in an array consisting of a number of conductors connected in series.

Nature of input impedance

6. (i) The distribution of the current maxima and minima is entirely independent of circuit adjustment at the transmitter. The tuning adjustments at this end may greatly alter the magnitude of the current, but will not affect the relative amplitudes or phases at any two points in the wire. The impedance offered to the transmitter will, however, vary greatly with the distance of the supply point from the far end. In fig. 2a the transmitter is connected at a distance of $\frac{\lambda}{4}$, and in fig. 2b, a distance of $\frac{3}{4}\lambda$, from the far end. In both cases the supply is connected at

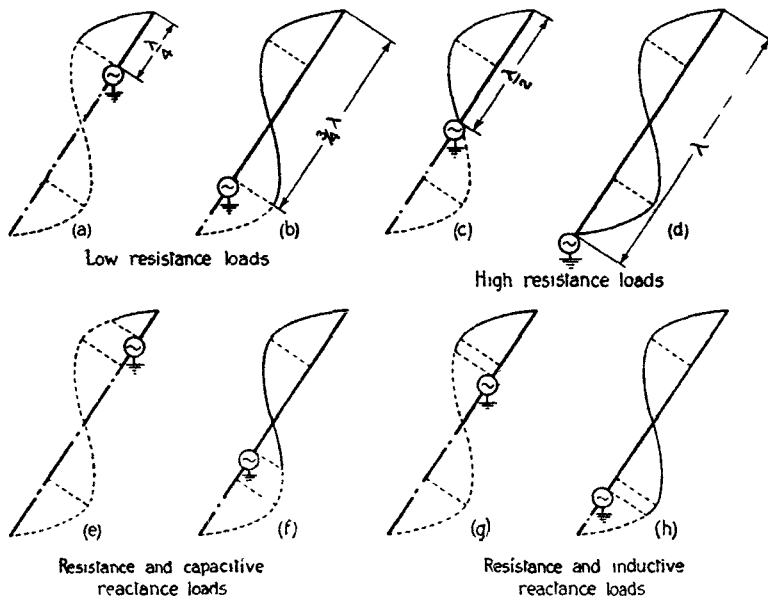


FIG. 2, CHAP. XV.—Nature of impedance of wires of various lengths

CHAPTER XV.—PARAS. 7-8

a current maximum, and whenever this is so, the wire offers an approximately non-reactive impedance of the order of 35 to 100 ohms. In fig. 2c, however, the transmitter is connected at a distance of $\frac{\lambda}{2}$, and in fig. 2d, a distance of λ , from the far end. In both cases the supply is connected at a current minimum and the wire offers an approximately non-reactive impedance of the order of 2,000 to 8,000 ohms.

(ii) Now suppose the generator to be connected at some intermediate point. In order to be quite clear it is preferable to draw a considerable portion of line, insert the theoretical current distribution and afterwards insert the generator. For example in figs. 2e and 2f, the generator has been inserted at a point less than $\frac{\lambda}{4}$ after a current *minimum*, measured from the far end. Under these conditions the impedance offered by the wire is equivalent to that of a condenser and resistance in series. In fig. 2g the generator has been inserted at a point between $\frac{\lambda}{4}$ and $\frac{\lambda}{2}$ and in fig. 2h, at a point between $\frac{3}{4}\lambda$ and λ , from the far end. In both these cases, the supply is connected at a point less than $\frac{\lambda}{4}$ after a current *maximum* and the input impedance of the wire is equivalent to that of an inductance and resistance in series.

7. (i) Theoretically, a wire in free space has zero reactance whenever its length is a multiple of $\frac{\lambda}{4}$. Actually, the velocity of wave propagation along copper wires located near to a conductive plane is about four per cent. less than the velocity of electro-magnetic waves in free space, and if a length of wire is to be non-reactive its length should be only about 0.96 of the theoretical value. Thus an aerial having an electrical length of $\frac{\lambda}{4}$ should be about 0.24 λ in actual length and so on. The exact location with respect to the ground, and the variation in permittivity and conductivity of the latter, must necessitate a slight variation of this figure in certain instances. In practice it may be found necessary to reduce the lengths of radiating members as much as ten per cent. below their nominal length, because discontinuities such as sharp bends, suspension insulators, etc., all tend to reduce the velocity of propagation along the wire.

(ii) In the early days of radio communication the frequencies employed were very much lower than those now in general use, and it was as a rule, only possible to employ aërials having a length very much less than $\frac{\lambda}{4}$. In these circumstances, the most important electrical constants of the aerial are (a) its effective resistance and (b) its effective capacitance. With the higher frequencies now in use, however, the input impedance may be equivalent to that of an inductive or capacitive resistance, or purely resistive, depending upon the ratio of length to λ .

Radiation resistance

8. The radiation resistance of an aerial is defined in Chapter VII; expressions giving the radiation resistance in certain simple cases are also contained therein. The conventional method of finding the radiation resistance of any given aerial is to develop an equation giving the field strength at all points in space. In Chapter I it is shown that the energy density of a uniform electric field of

strength E is $\frac{\kappa E^2}{8\pi}$ ergs per cubic centimetre. If the calculated value of field strength is inserted in this expression, and the result multiplied by the velocity of propagation, we obtain the energy per second, i.e. the power, which is passing through any unit area in a plane perpendicular to

the direction of propagation. The total power passing through a sphere surrounding the radiator is obviously equal to that radiated, hence, on summing up the total power passing through every unit area on the surface of this sphere, and then dividing by the square of the R.M.S. loop current, the quotient is the radiation resistance. Another method is to sum up the energy passing through a cylinder of unit thickness immediately surrounding the wire. This method gives both the radiated power and the wattless volt-amperes required to maintain the induction field, and therefore gives the aerial impedance as the vector cum of the radiation resistance and the reactance of the aerial. In this manner the impedance of a $\frac{\lambda}{2}$ dipole in free space

is found to be $73.3 + j42.5$ ohms, and the impedance of a verticle $\frac{\lambda}{4}$ aerial over a perfectly conductive earth, $36.6 + j21.25$ ohms. When tuned to resonance with the frequency of the supply, the reactance of the aerial is annulled, although of course the induction field is still maintained.

9. (i) In practice, the radiation resistance is affected by the proximity of the ground, to an extent depending upon the permittivity and conductivity of the soil. The radiation resistance of a vertical $\frac{\lambda}{2}$ dipole, with its lower end at ground level on a perfectly conductive earth, is approximately 100 ohms, whereas in free space it would be 73.3 ohms. The nature of the variation with height above ground level is shown in fig. 3. Over moist earth of permittivity $\epsilon = 25$ and conductivity $\sigma = 10^8$ E.S.U. the actual radiation resistance is found to be very close to the theoretical value given by this curve, which may therefore be used for practical purposes. Application of the image theorem of paragraph 35 shows that a horizontal $\frac{\lambda}{2}$ dipole very close indeed to the earth's surface will radiate infinitesimal energy, i.e. the radiation resistance of such an aerial approaches zero. When far above the surface, however, its radiation resistance is 73.3 ohms, the theoretical nature of the variation with height being also shown in fig. 3. Over the ground specified above, however,

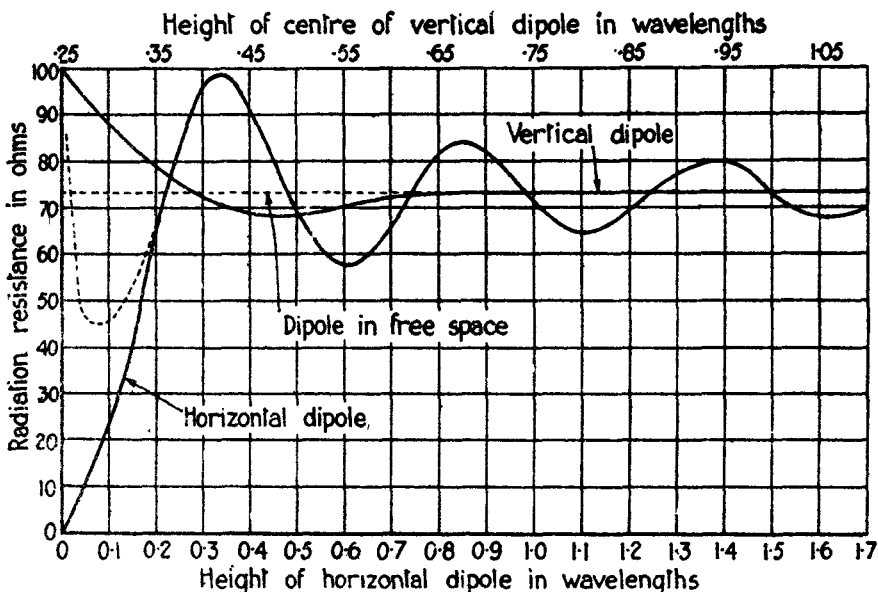


FIG. 3, CHAP. XV.—Radiation resistance of dipole.

CHAPTER XV.—PARA. 10

the resistance at heights less than $\frac{\lambda}{4}$ was found experimentally to be given by the curve shown in dotted line. It will be observed that in both solid-line curves, the radiation resistance approaches its free-space value in an oscillating manner.

(ii) The radiation resistance of an aerial array can be calculated by either of the methods previously referred to, but except for very simple arrays, the labour is prohibitive. In any event, it is impossible to define the radiation resistance of an array of which the various members carry currents of different magnitudes, except by the somewhat arbitrary method of referring it to the current in some particular member. Subject to this limitation, however, it is possible to compute the radiation resistance of a simple array from that of each member and the mutual impedance between the various members.

Radiation from hertzian doublet

10. (i) As an introduction to the theory of aerial arrays we may first consider the radiation field, γ_0 , in the equatorial plane of a single hertzian doublet of length l , situated in free space, fig. 4a. In this theoretical radiator, a conduction current $i = \mathcal{I} \cos \omega t$ is assumed to flow in the wire joining the two capacitance areas, and a corresponding displacement current in the dielectric between them, so that the elementary portions of the conductor itself are supposed to have no capacitance, so that the amplitude $\mathcal{I}$ of the current is the same at all points in the wire. The conduction current consists of a number N of electrons of charge e E.S.U. = q coulombs, the average instantaneous velocity being, say, b centimetres per second. Let the cross-section of the conductor be A cm²; the total volume of wire is then Al cm³ and the density of the moving

charge inside the wire is $\frac{Nq}{Al}$ coulombs per cm³. The current density will therefore be $\frac{Nq}{Al} \cdot \frac{\text{coulomb}}{\text{cm}^3} \times b \frac{\text{cm}}{\text{sec}} = \frac{Nqb}{Al}$ coulomb per second per cm³ or amperes per cm³, and the

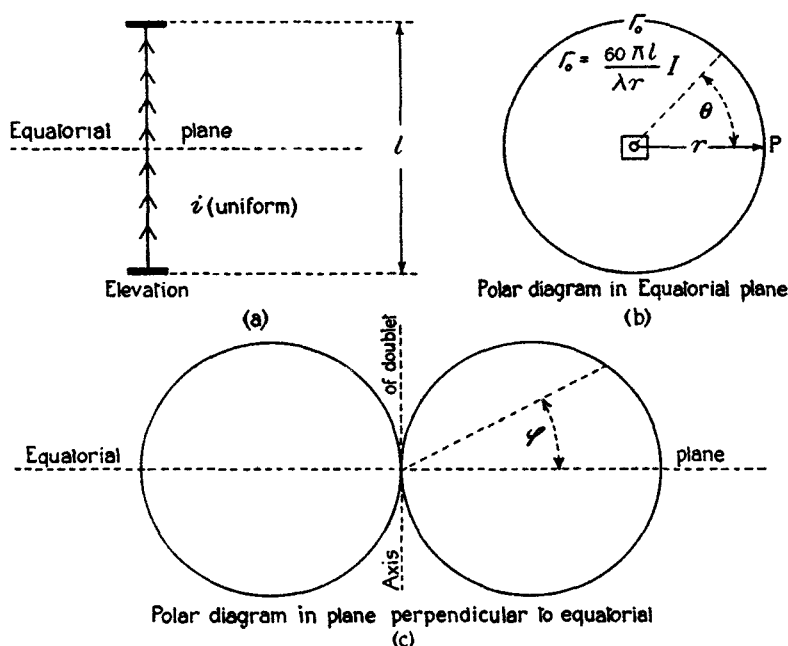


FIG. 4, CHAP. XV.—Radiation from hertzian doublet.

total instantaneous current $\frac{Nqb}{l}$ amperes. We may therefore write

$$i = \frac{Nqb}{l} = \mathcal{I} \cos \omega t,$$

$$\therefore b = \frac{l \mathcal{I}}{Nq} \cos \omega t. \quad \dots \dots \dots (1)$$

The radiation depends upon the acceleration $a = \frac{db}{dt}$ of the electrons. Since b is of sinusoidal form,

$$a = -\frac{\omega l \mathcal{I}}{Nq} \sin \omega t$$

$$= \frac{\omega l \mathcal{I}}{Nq} \cos \left(\omega t + \frac{\pi}{2} \right). \quad \dots \dots \dots (2)$$

(ii) The object of keeping the above expression in cosine form is to bring into prominence the relative phase of the radiation and the current. In Chapter VII it is shown that the electric field strength γ due to a single accelerated electron at a point P, distant r centimetres from the

centre O of the doublet, and in the equatorial plane, is equal to $\frac{ae}{c^2 r}$ dynes per unit charge. If then

the electrons in the wire are sinusoidally accelerated, the field at the point P is also sinusoidal, but will lag behind the acceleration producing it by an angle δ , owing to the finite velocity of propagation, c . Instead of expressing the field strength in dynes per unit charge, it is more convenient to express it in practical units. The field set up by the oscillating charge of q coulombs

is therefore given by substituting, in the formula $\gamma = \frac{ae}{c^2 r}$, $\frac{9 \times 10^{11} \times \omega l \mathcal{I}}{Nq} \cos \left(\omega t + \frac{\pi}{2} \right)$ for the acceleration a , and Nq for the charge e , giving

$$\gamma_0 = 9 \times 10^{11} \omega l \mathcal{I} \cos \left(\omega t + \frac{\pi}{2} - \delta \right). \quad \dots \dots (3)$$

It is now convenient to put $c = 3 \times 10^{10}$, $\omega = 2\pi f = \frac{2\pi c}{\lambda}$, giving the amplitude $\hat{\gamma}_0$ as

$$\hat{\gamma}_0 = 9 \times 10^{11} \times \frac{2\pi c}{\lambda} \times \frac{l \mathcal{I}}{c^2 r}$$

$$= \frac{60 \pi l}{\lambda r} \mathcal{I}.$$

Thus the complete expression for the electric field in volts per centimetre is

$$\gamma_0 = \frac{60 \pi l}{\lambda r} \mathcal{I} \cos \left(\omega t + \frac{\pi}{2} - \delta \right). \quad \dots \dots \dots (4)$$

Note that the lengths l , λ , and r are all measured in centimetres. If these are given in metres the field strength is expressed in volts per metre; or if the constant 60 is replaced by 37·25, and r is measured in miles (l and λ still in metres) the field strength is in millivolts per metre.

11. The amplitude of the field is seen to vary inversely as the distance r , but is independent of the angle θ . Hence the polar diagram in the equatorial plane is a circle with the axis of the

CHAPTER XV.—PARA. 12

dipole as centre, fig. 4b. The phase angle δ obviously depends upon the distance r , for the wave travels this distance in a time $\frac{r}{c}$, hence $\delta = \frac{\omega r}{c} = \frac{2\pi}{\lambda} r$, and therefore

$$\gamma_0 = \frac{60 \pi l}{\lambda r} \mathcal{J} \cos \left(\omega t + \frac{\pi}{2} - \frac{2\pi}{\lambda} r \right). \quad \dots \dots \dots (5)$$

If the point P, instead of being in the equatorial plane, is situated at an angle φ above it, the field strength will be proportional to $\cos \varphi$ and is

$$\gamma_\varphi = \frac{60 \pi l}{\lambda r} \mathcal{J} \cos \varphi \cos \left(\omega t + \frac{\pi}{2} - \frac{2\pi}{\lambda} r \right). \quad \dots \dots \dots (6)$$

Considering the amplitude only, we see that this varies with the angle φ ; in the equatorial plane, $\varphi = 0$, and the amplitude of the field strength is $\frac{60 \pi l}{\lambda r} \mathcal{J}$, while in the polar direction, $\varphi = 90^\circ$, the amplitude is zero. The variation of γ with φ is easily shown by means of a polar diagram as in fig. 4c in which the radius vector is proportional to $\cos \varphi$. The diagram therefore consists of two circles of unit diameter which are in contact at the axis of the doublet. By rotating the diagram round this axis we obtain a solid surface giving the relative field strengths in all directions in space. It is most important to remember that this solid surface does not represent the wave front, the latter being a spherical surface.

Radiation from half-wave dipole

12 The half-wave aerial differs from the hertzian doublet in its current distribution. Let the current at the centre of the dipole be $i = \mathcal{J}_0 \cos \omega t$. If the distance, $y = OA$, fig. 5, is measured from an origin O at the centre of the wire, the peak current $\mathcal{J}$, at the point A is $\mathcal{J}_0 \cos \frac{2\pi y}{\lambda}$. To find the field strength at a point P, distant r from the origin and in the equatorial plane, we consider the field set up by an elementary length dy of conductor, distant

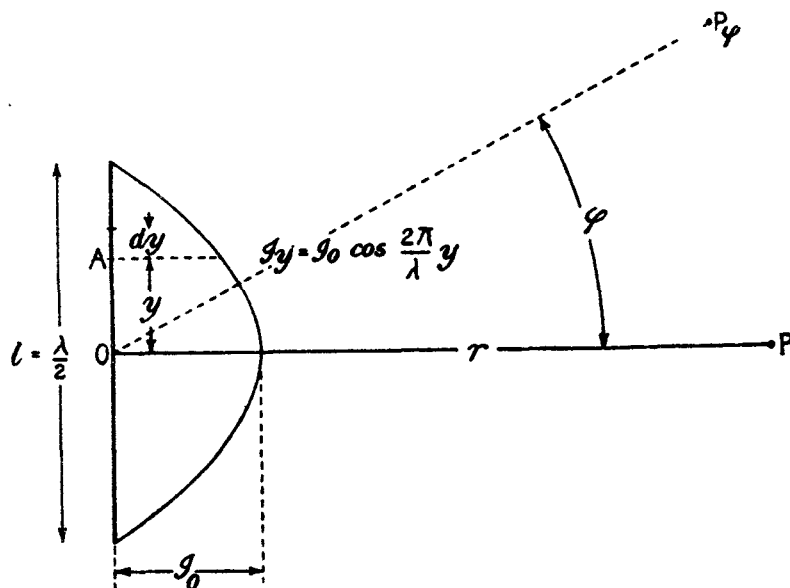


FIG. 5, CHAP. XV.—Notation for dipole in free space.

y from the origin ; the current over this short length is practically uniform and we may therefore treat the element of conductor as a hertzian doublet of length dy . Obviously the field $d\gamma_0$ at P, due to this element of conductor, is

$$d\gamma_0 = \left(\frac{60\pi}{\lambda r} \mathcal{J}_0 \cos \frac{2\pi y}{\lambda} \right) \cos \left(\omega t + \frac{\pi}{2} - \frac{2\pi}{\lambda} r \right).$$

Now the fields produced at P by all the elements of conductor will be in phase, and the total field is that contributed by all such elements above and below the origin, i.e. from $y = +\frac{\lambda}{4}$ to $y = -\frac{\lambda}{4}$, and

$$\begin{aligned} \gamma_0 &= \frac{60\pi}{\lambda r} \mathcal{J}_0 \cos \left(\omega t + \frac{\pi}{2} - \frac{2\pi}{\lambda} r \right) \int_{y=-\frac{\lambda}{4}}^{y=+\frac{\lambda}{4}} \cos \frac{2\pi y}{\lambda} \cdot dy \\ &= \left[\frac{60\pi}{\lambda r} \mathcal{J}_0 \cos \left(\omega t + \frac{\pi}{2} - \frac{2\pi}{\lambda} r \right) \right] \times \frac{\lambda}{\pi} \\ &= \frac{60}{r} \mathcal{J}_0 \cos \left(\omega t + \frac{\pi}{2} - \frac{2\pi}{\lambda} r \right). \quad \dots \dots \dots (7) \end{aligned}$$

At a point P, above or below the equatorial plane, subtending with the latter an angle φ , the field is not now merely proportional to $\cos \varphi$ as in the case of the hertzian doublet. Instead, a

factor $\frac{\cos \left(\frac{\pi}{2} \sin \varphi \right)}{\cos \varphi}$ must be introduced, giving

$$\gamma_\varphi = \frac{60}{r} \frac{\cos \left(\frac{\pi}{2} \sin \varphi \right)}{\cos \varphi} \mathcal{J}_0 \cos \left(\omega t + \frac{\pi}{2} - \frac{2\pi}{\lambda} r \right). \quad \dots \dots \dots (8)$$

The factor $\frac{\cos \left(\frac{\pi}{2} \sin \varphi \right)}{\cos \varphi}$ is plotted in fig. 6. It is seen to consist of two approximately elliptical figures, in contact on the axis of the dipole. By rotating this figure we obtain the solid polar diagram as in the previous instance.

13. Collecting the principal formulæ so far developed, we have the following expressions for the amplitude $\hat{r}_\varphi$ of the electric field at an angle φ with respect to the equatorial plane.

(i) For the hertzian doublet

$$\hat{r}_\varphi = \frac{60 \pi l}{\lambda r} \mathcal{J}_0 \cos \varphi,$$

(ii) For the $\frac{\lambda}{2}$ dipole

$$\hat{r}_\varphi = \frac{60}{r} \mathcal{J}_0 \frac{\cos \left(\frac{\pi}{2} \sin \varphi \right)}{\cos \varphi}.$$

CHAPTER XV.—PARA. 14

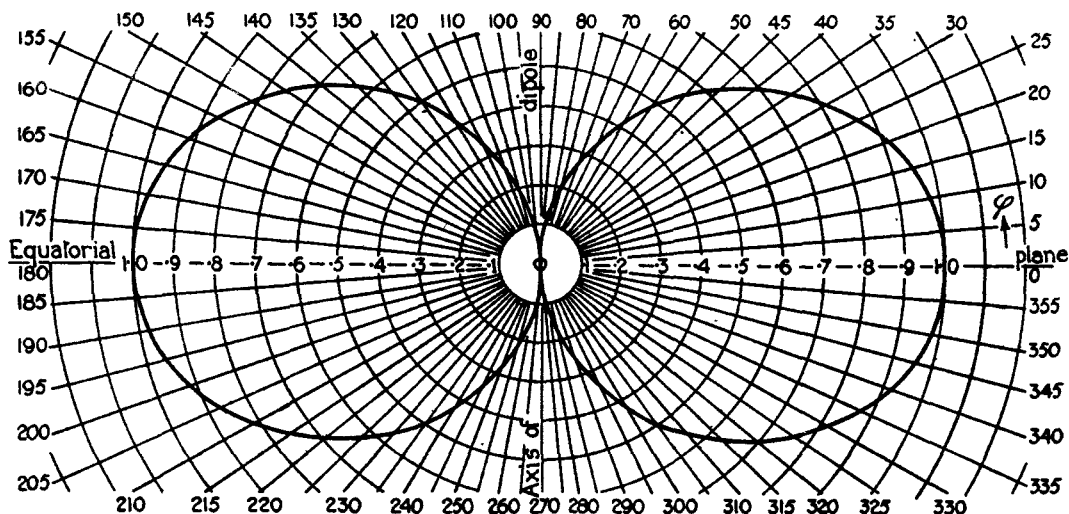


FIG. 6, CHAP. XV.—Current Distribution Factor for $\frac{\lambda}{2}$ dipole.

Each of these expressions may be divided into three factors,

(a) $\frac{60}{r} \mathcal{J}_0$, which is common to both.

(b) In (i), $\frac{\pi l}{\lambda}$. The corresponding factor in (ii) is unity. These factors depend upon the distribution of capacitance along the aerial, and may be called the respective Form Factors. Thus the Form Factor of a hertzian doublet is $\frac{\pi l}{\lambda}$ and of a $\frac{\lambda}{2}$ dipole is unity.

(c) In (i), $\cos \varphi$. The corresponding factor in (ii) is $\frac{\cos \left(\frac{\pi}{2} \sin \varphi \right)}{\cos \varphi}$.

These factors take into account the effect of the current distribution upon the relative phases of the elements of field strength at points above and below the equatorial plane, and may be termed the respective Current Distribution Factors. In general, if the Form Factor is denoted by F , and the Current Distribution Factor by $f(\varphi)$, the amplitude of the electric field at a distance r may be written

$$\hat{E}_\varphi = \frac{60}{r} \mathcal{J}_0 \cdot F \cdot f(\varphi).$$

Power input

14. Suppose it is required to produce a certain field strength at a point P, distant r miles from the radiator, where r is sufficiently great to justify the neglect of the induction field but so near that the effects of attenuation are negligible. Working in R.M.S. values, we have from the previous discussion

$$E = \frac{37 \cdot 25}{r} F \cdot f(\varphi) I \text{ (millivolts/metre)}$$

and if the radiation resistance of the aerial is R_r ohms the power radiated is P_r watts, where

$$P_r = I^2 R_r.$$

If P_r is given, then, the required current is

$$I = \sqrt{\frac{P_r}{R_r}}$$

and the field strength at a distance r is

$$E = \frac{37 \cdot 25}{r} F \cdot f(\varphi) \sqrt{\frac{P_r}{R_r}}$$

$$P_r = \left(\frac{r E}{37 \cdot 25 \times F \times f(\varphi)} \right)^2 R_r.$$

Also, the input power will be given by

$$P_i = \left(\frac{r E}{37 \cdot 25 \times F \times f(\varphi)} \right)^2 R_A$$

where R_A is the total resistance of the aerial. For example let the point P be in the equatorial plane of a $\frac{\lambda}{2}$ dipole. Then $F = 1$, $f(\varphi) = 1$. Let the radiation resistance be 73 ohms and the loss resistance 10 ohms. Then to produce a field of 100 millivolts per metre at a distance of one mile, the power input must be

$$P_i = \left(\frac{1 \times 100}{37 \cdot 25} \right)^2 \times 83$$

$$= 593 \text{ watts.}$$

The power actually radiated will be $\frac{73}{83}$ of this or 522 watts

Field due to two parallel dipoles

15. Let us now consider the field strength produced by a simple array consisting of two parallel half-wave dipoles in free space; these are spaced apart by a distance d as shown in plan and elevation in fig. 7, where A and B are the wires and O a point which will be regarded as an origin. We will first calculate the radiation in the equatorial plane perpendicular to the wires, each of which is assumed to carry a current $i = \mathcal{I} \cos \omega t$, i.e. the currents in the two wires are in phase. As before, consider the field at a point P, distant r from the origin O, and let $r \gg \lambda$. Then AP, OP and BP are practically parallel to each other, and $OP = r$, $AP = r + \frac{d}{2} \cos \theta$, $BP = r - \frac{d}{2} \cos \theta$.

The field produced by the current in the wire A will be

$$E_A = K \mathcal{I}_0 \cos \left\{ \omega t + \frac{\pi}{2} - \frac{2\pi}{\lambda} \left(r + \frac{d}{2} \cos \theta \right) \right\} \quad \dots \quad (9a)$$

where $K = \frac{60}{r + \frac{d}{2} \cos \theta} + \frac{60}{r}$; similarly the current in B will produce a field

$$E_B = K \mathcal{I}_0 \cos \left\{ \omega t + \frac{\pi}{2} - \frac{2\pi}{\lambda} \left(r - \frac{d}{2} \cos \theta \right) \right\} \quad \dots \quad (9b)$$

CHAPTER XV.—PARA. 16

and the total field $\gamma_o = \gamma_A + \gamma_B$ or

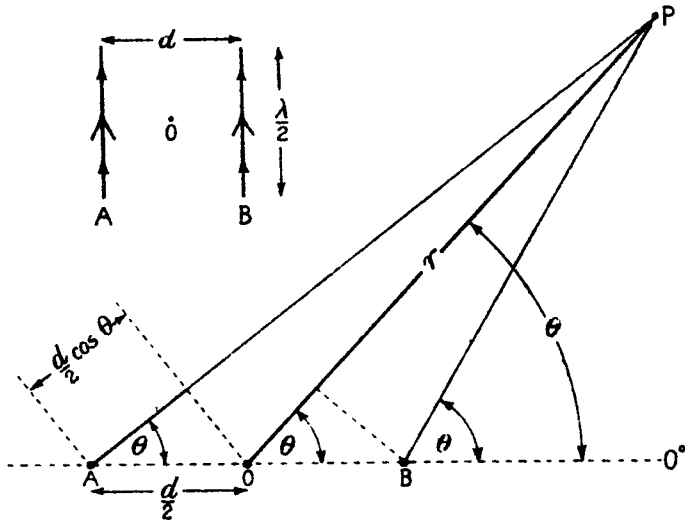
$$\gamma_o = K \mathcal{J}_o \left[\cos \left\{ \omega t + \frac{\pi}{2} - \frac{2\pi}{\lambda} r - \frac{\pi d}{\lambda} \cos \theta \right\} + \cos \left\{ \omega t + \frac{\pi}{2} - \frac{2\pi}{\lambda} r + \frac{\pi d}{\lambda} \cos \theta \right\} \right]. \quad \dots \quad (9c)$$

By a formula developed in Chapter V, this may be written

$$\gamma_o = 2K \mathcal{J}_o \cos \left(\frac{\pi d}{\lambda} \cos \theta \right) \left\{ \cos \left(\omega t + \frac{\pi}{2} - \frac{2\pi}{\lambda} r \right) \right\}. \quad \dots \quad (10)$$

Thus the amplitude of the electric field is

$$\hat{r}_o = \frac{60}{r} \mathcal{J}_o \times 2 \cos \left(\frac{\pi d}{\lambda} \cos \theta \right).$$



NB As $OP \gg AB$, AP , OP , BP are considered to be parallel

FIG. 7, CHAP. XV.—Notation for parallel dipoles in free space.

Grating factor

16. The amplitude of the field due to a single dipole situated at the origin O would be

$$\hat{r}_o = \frac{60}{r} \mathcal{J}_o$$

and we see that the field due to two parallel dipoles, separated by a distance d , and carrying equal, syn-phased currents, is obtained by multiplying that of a single dipole by the factor

$2 \cos \left(\frac{\pi d}{\lambda} \cos \theta \right)$, which is called the equatorial plane Grating Factor for a pair of dipoles. Its

value obviously lies between the limits zero and 2, and is plotted in polar co-ordinates in fig. 8,

line A, for various values of $\frac{d}{\lambda}$ from 0 to 4. The circle surrounding each diagram has a radius

of 2 units, representing the upper limiting value of the Grating Factor. The first diagram, fig. 8, A 1, is for $d = 0$, i.e. two superimposed dipoles each carrying unit current, which are

equivalent to a single dipole carrying a current of two units, hence the diagram corresponds with the limiting circle. For other values of $\frac{d}{\lambda}$, upon the line joining the two aerials, the field strength varies with the spacing. As d is increased toward the value $\frac{\lambda}{2}$, the field strength gradually decreases, and when $d = \frac{\lambda}{2}$ the radiation from the two aerials is in anti-phase at all points along this line, so that complete interference, i.e. cancellation, results. For values of d greater than $\frac{\lambda}{2}$, multiple lobes appear. In the directions $\theta = 0$ and $\theta = 180^\circ$, the Grating Factor is 2, if n is even, and zero, if n is odd, whenever $d = \frac{n\lambda}{2}$. Upon a line perpendicular to that joining the two aerials, the Grating Factor is equal to 2 for all values of d , because the radiation from both aerials reaches all points simultaneously.

17. Now consider the field produced in the equatorial plane by two dipoles in which the currents are of equal magnitude but differ in phase. Let the aerials be A, carrying a current $i_A = \mathcal{I}_0 \cos \omega t$ and B with the current $i_B = \mathcal{I}_0 \cos (\omega t + \beta)$. Then

$$\begin{aligned} \gamma_A &= K \mathcal{I}_0 \cos \left(\omega t + \frac{\pi}{2} - \frac{2\pi}{\lambda} r - \frac{\pi d}{\lambda} \cos \theta \right) \\ \gamma_B &= K \mathcal{I}_0 \cos \left(\omega t + \frac{\pi}{2} - \frac{2\pi}{\lambda} r + \frac{\pi d}{\lambda} \cos \theta + \beta \right). \end{aligned}$$

$$\text{If } \left(\omega t + \frac{\pi}{2} - \frac{2\pi}{\lambda} r \right) = \omega t_0$$

$$\gamma_A + \gamma_B = K \mathcal{I}_0 \left\{ \cos \left(\omega t_0 - \frac{\pi d}{\lambda} \cos \theta \right) + \cos \left(\omega t_0 + \frac{\pi d}{\lambda} \cos \theta + \beta \right) \right\}$$

and the total field becomes

$$\gamma_0 = K \mathcal{I}_0 \left[\cos \left(\omega t_0 - \frac{\pi d}{\lambda} \cos \theta \right) + \cos \left(\omega t_0 + \frac{\pi d}{\lambda} \cos \theta + \beta \right) \right]. \quad (11)$$

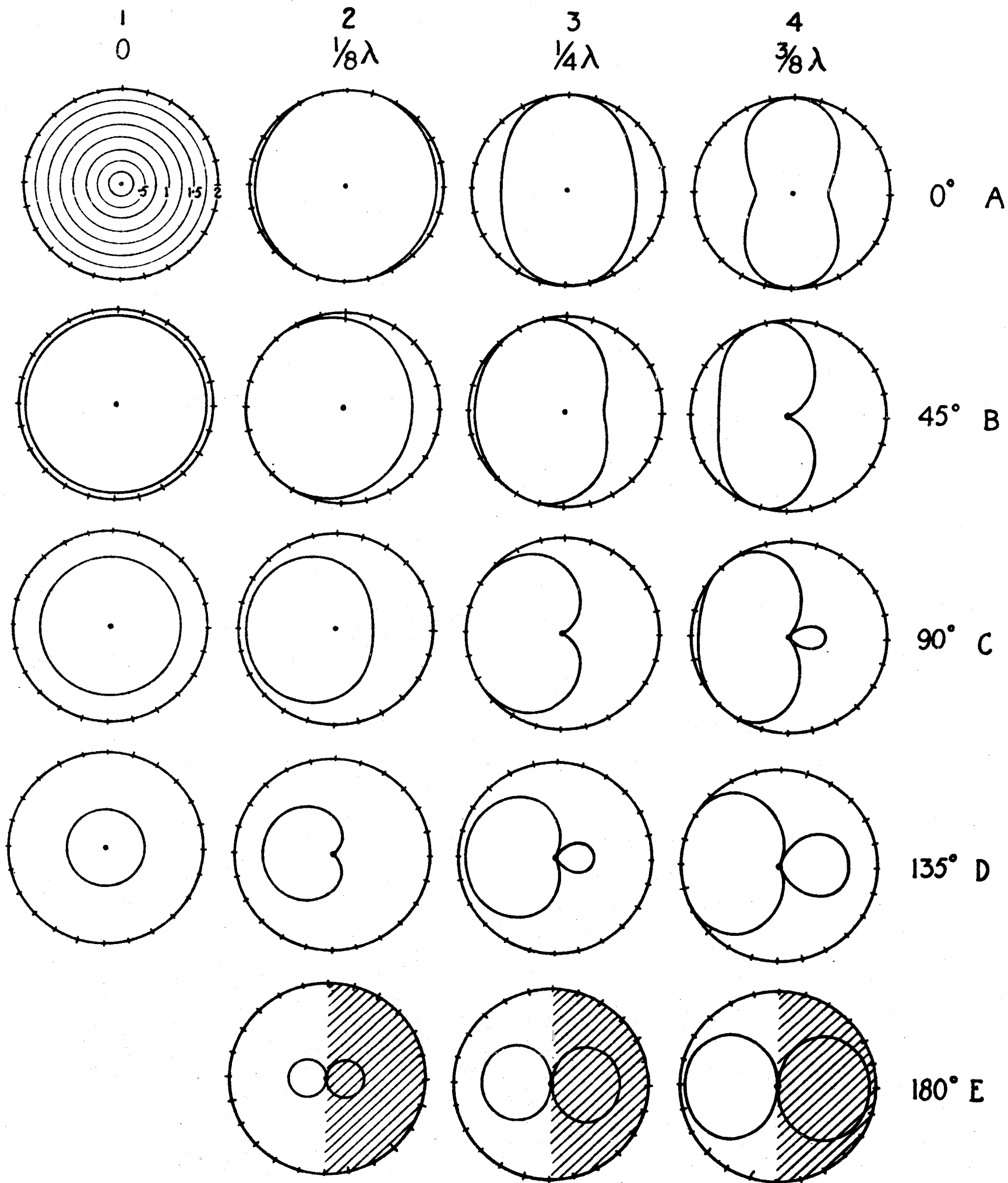
To simplify, put $\omega t_0 = X$, $\frac{\pi d}{\lambda} \cos \theta = Y$, $P = X - Y$, $R = X + Y$, $Q = R + \beta$; then

$$\begin{aligned} \gamma_0 &= K \mathcal{I}_0 \{ \cos (X - Y) + \cos (X + Y + \beta) \} \\ &= K \mathcal{I}_0 \{ \cos P + \cos Q \}. \end{aligned}$$

$$\text{By Chapter V, } \cos P + \cos Q = 2 \cos \frac{P+Q}{2} \cos \frac{P-Q}{2}.$$

Whence

$$\begin{aligned} \gamma_0 &= 2K \mathcal{I}_0 \cos \frac{X - Y + (X + Y + \beta)}{2} \cos \frac{(X - Y) - (X + Y) - \beta}{2} \\ &= 2K \mathcal{I}_0 \left\{ \cos \left(X + \frac{\beta}{2} \right) \cos \left(Y + \frac{\beta}{2} \right) \right\} \\ &= 2K \mathcal{I}_0 \left[\cos \left(\omega t_0 + \frac{\beta}{2} \right) \right] \cos \left(\frac{\pi d}{\lambda} \cos \theta + \frac{\beta}{2} \right). \quad \dots \quad (12) \end{aligned}$$



GRATING FACTOR (SHEET 1)

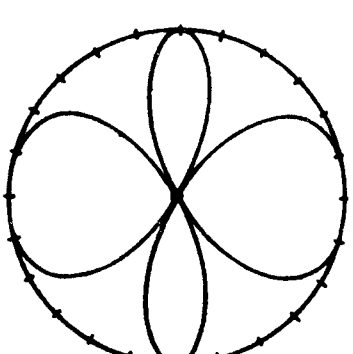
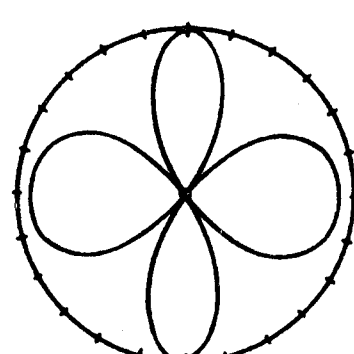
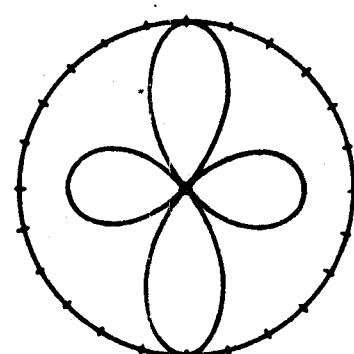
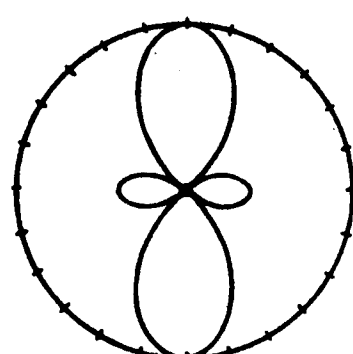
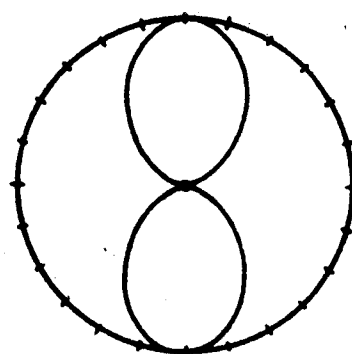
$\frac{5}{2} \lambda$

$\frac{6}{8} \lambda$

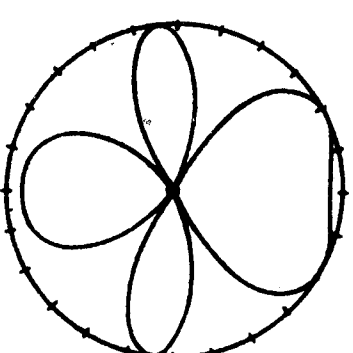
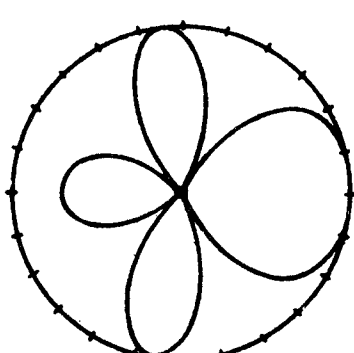
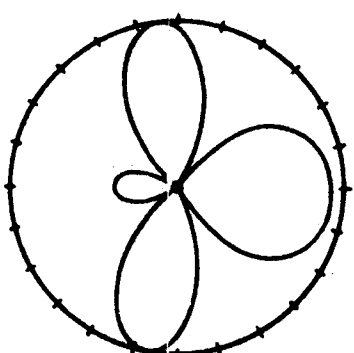
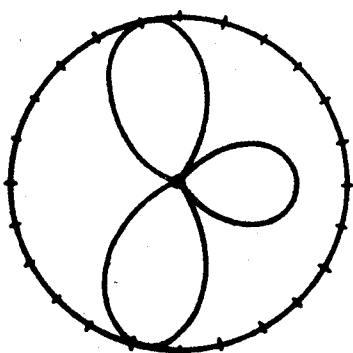
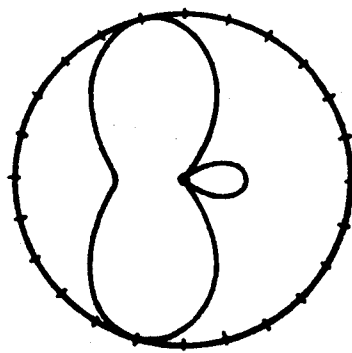
$\frac{7}{4} \lambda$

$\frac{8}{8} \lambda$

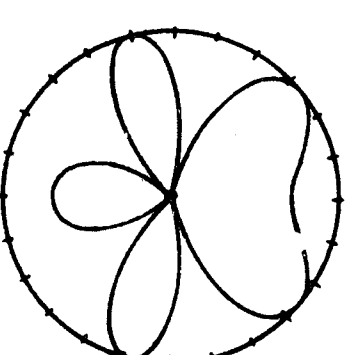
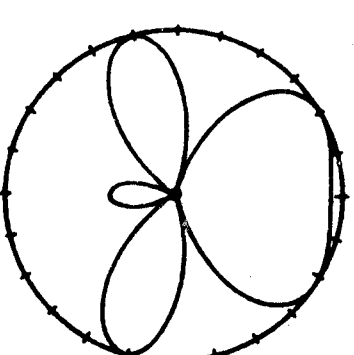
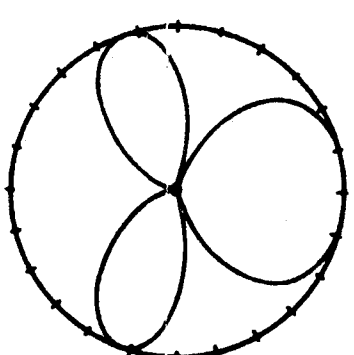
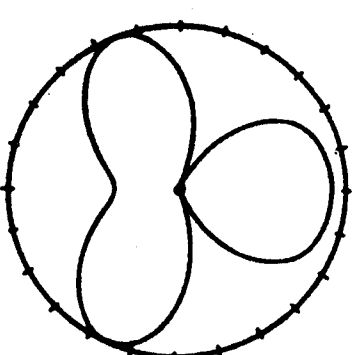
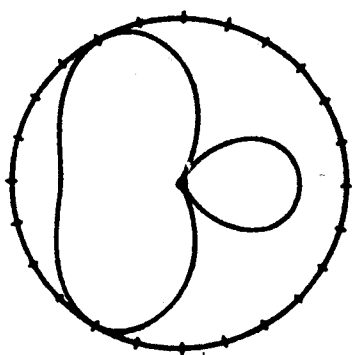
λ



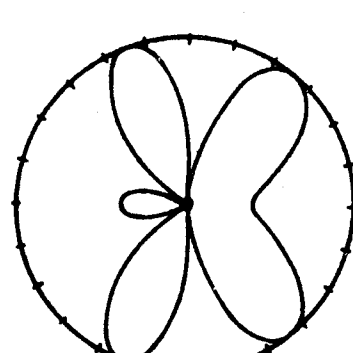
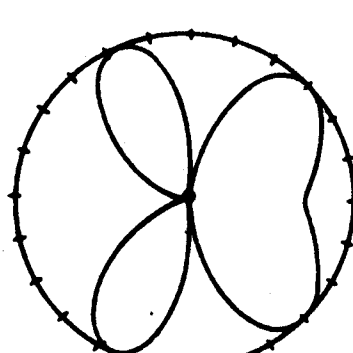
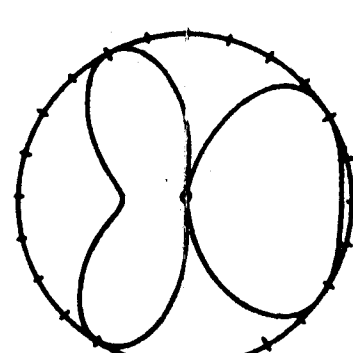
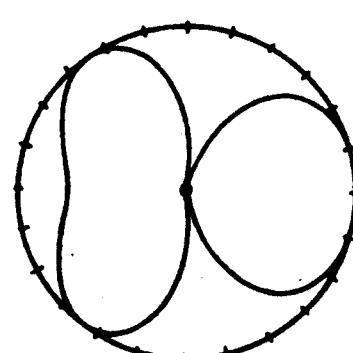
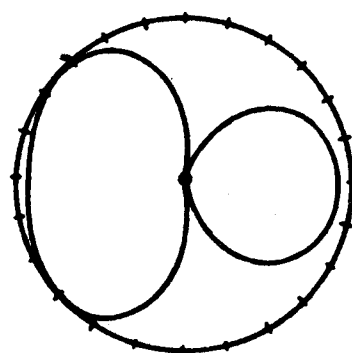
0° A



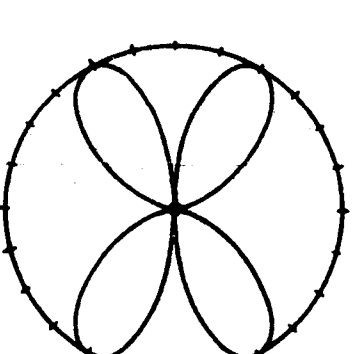
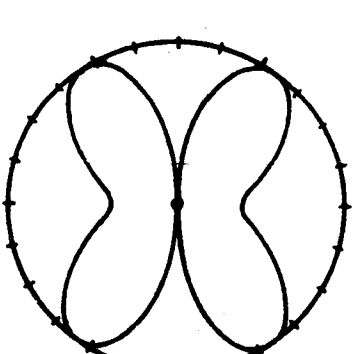
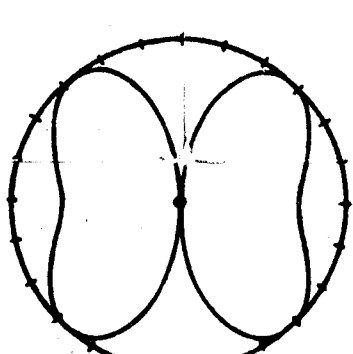
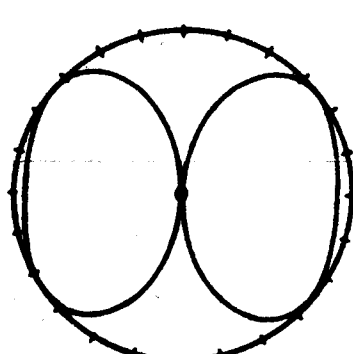
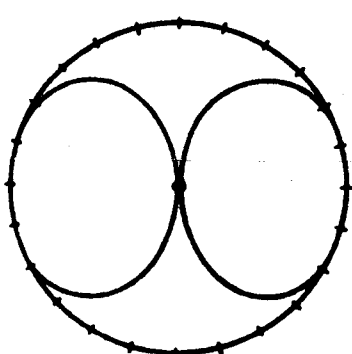
45° B



90° C



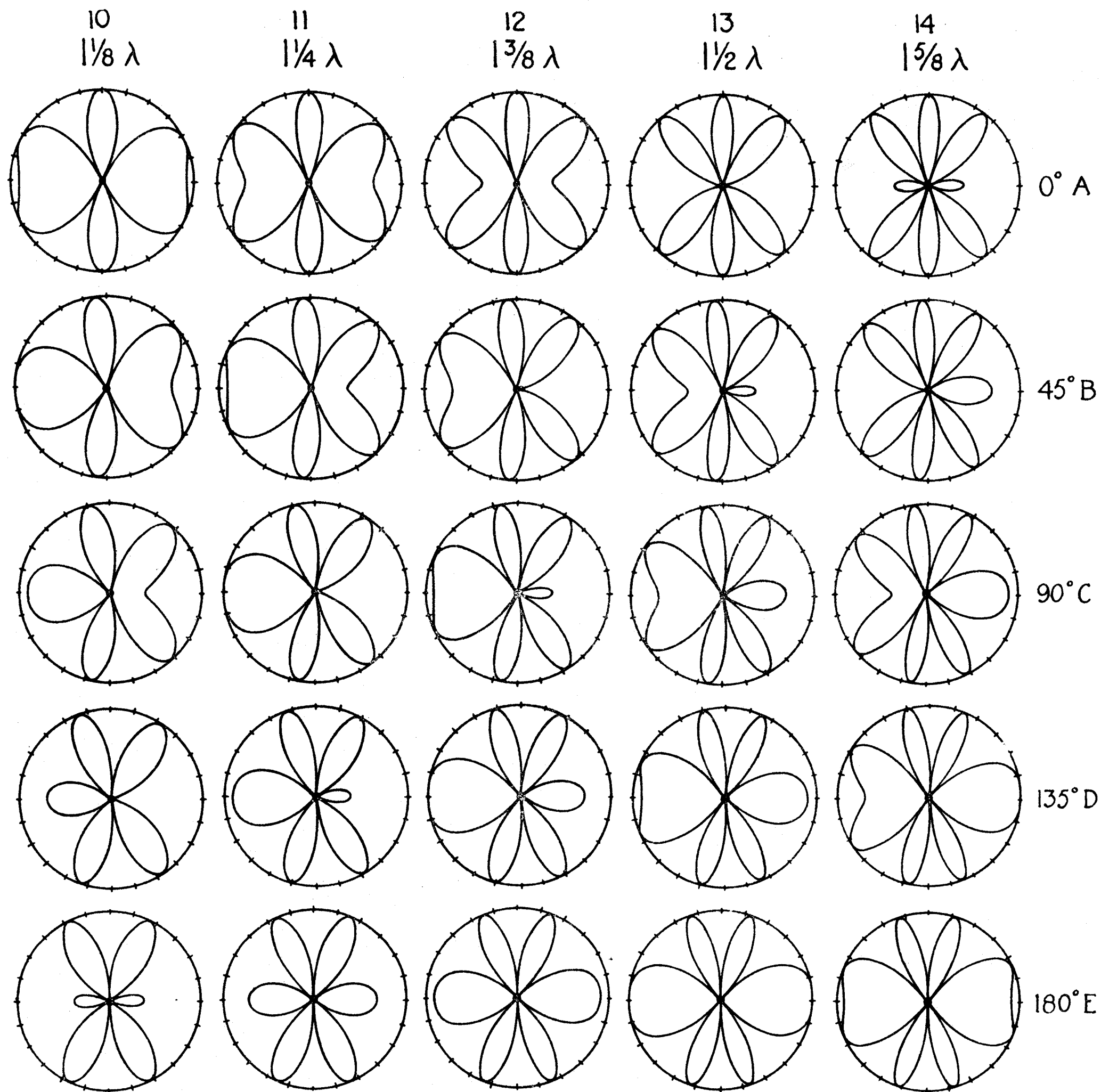
135° D



180° E

GRATING FACTOR (SHEET 2)

FIG. 8,
CHAP. XV



GRATING FACTOR (SHEET 3)

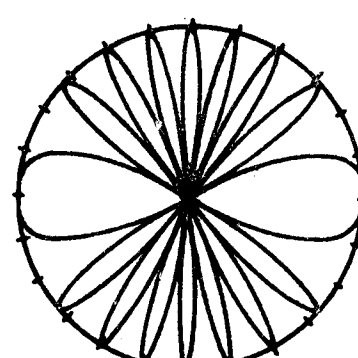
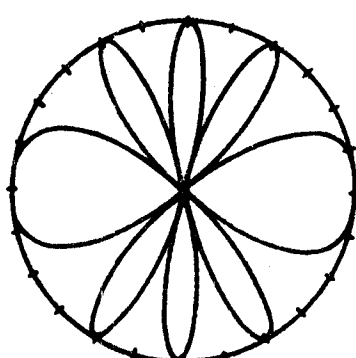
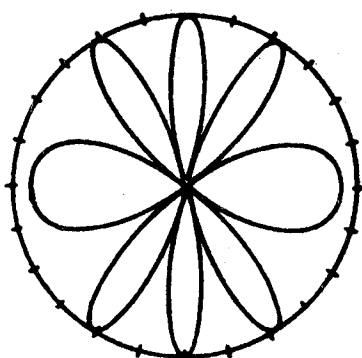
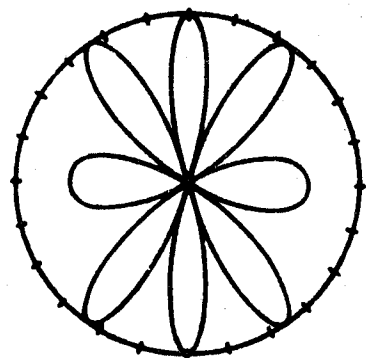
FIG. 8,
CHAP. XV

15
 $1\frac{3}{4}\lambda$

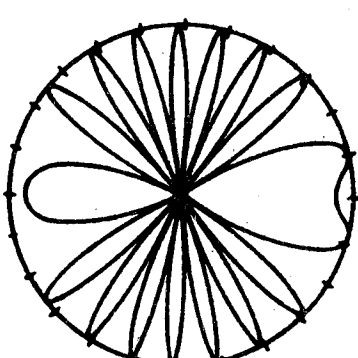
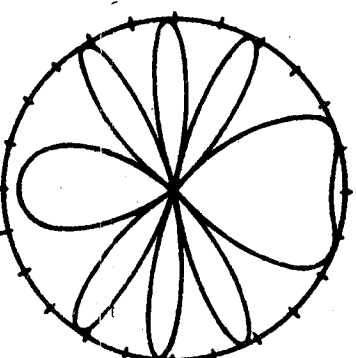
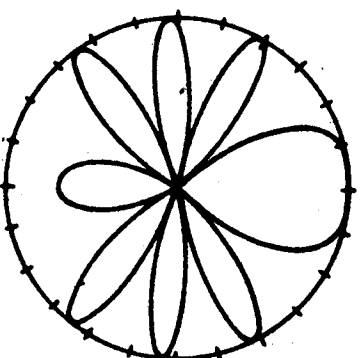
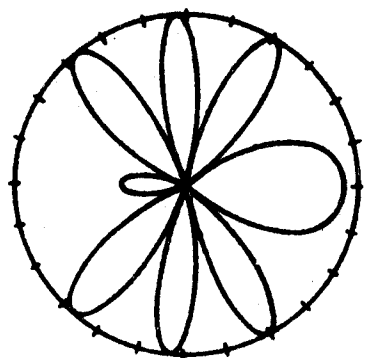
16
 $1\frac{7}{8}\lambda$

17
 2λ

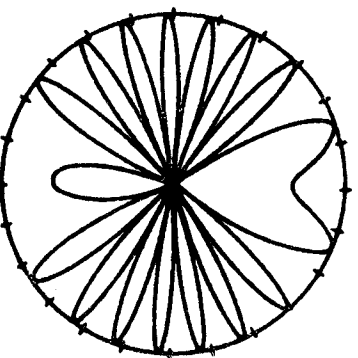
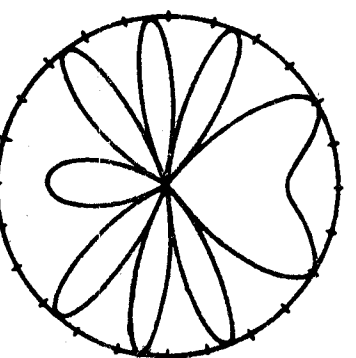
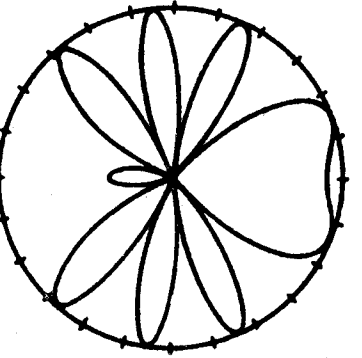
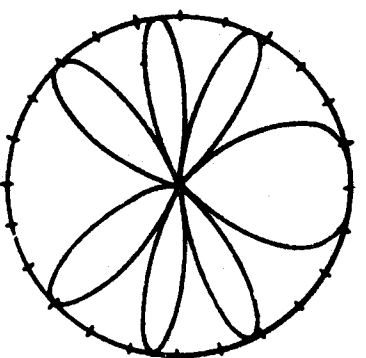
18
 4λ



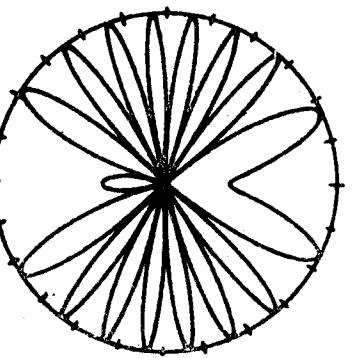
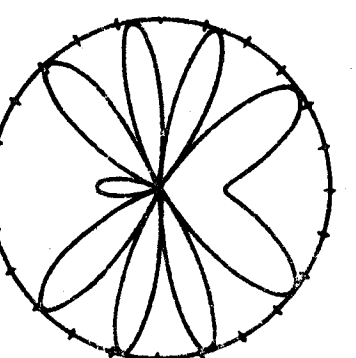
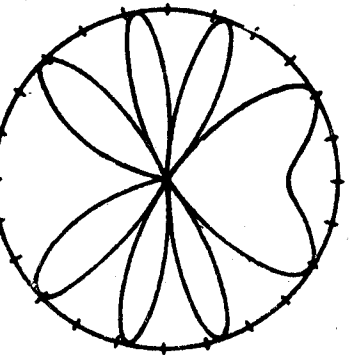
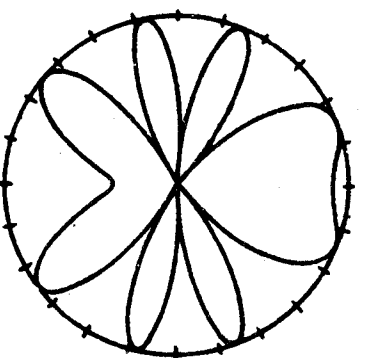
$0^\circ A$



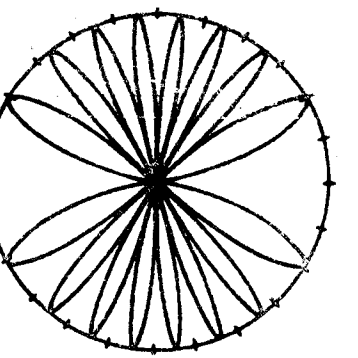
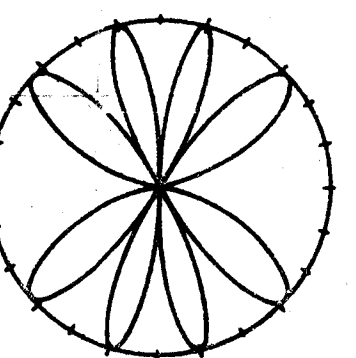
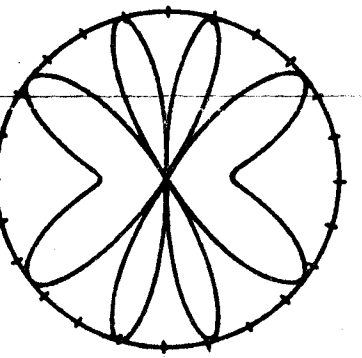
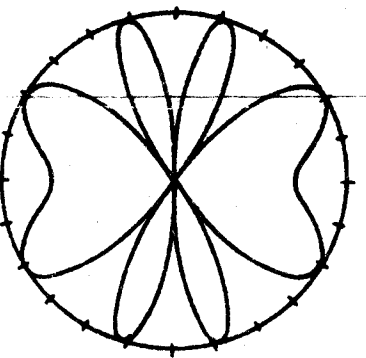
$45^\circ B$



$90^\circ C$



$135^\circ D$



$180^\circ E$

GRATING FACTOR (SHEET 4)

FIG. 8;
CHAP. XV

CHAPTER XV.—PARA. 18

The Grating Factor is therefore $2 \cos \left(\frac{\pi d}{\lambda} \cos \theta + \frac{\beta}{2} \right)$ and is plotted in polar co-ordinates in fig. 8, lines B to E, for various values of β and $\frac{d}{\lambda}$. Of particular interest is the bottom row, which shows the fields due to two aerials carrying currents in anti-phase. Obviously, the radiation from the two cancels out along a line perpendicular to that joining the two aerials, while along that line, the Grating Factor is zero if d is an even number of half wavelengths, and equal to 2, if d is an odd number of half wavelengths. Line C will again be referred to in connection with reflector aerials, while line E is of importance in the study of loop aerials, both for reception and transmission. We see then that fig. 8 has many important applications and will repay a very careful study. To facilitate the enlarged reproduction of any particular diagram, each limiting circle has been divided at 15° intervals, and a series of concentric circles of various radii inserted within the limiting circle of fig. 8, A 1. When adding these diagrams, it must be noted that the radius vector changes sign on passing through a zero. An example is given in paragraph 41.

Combinations of pairs of dipoles

18. Consider an array of four parallel dipoles spaced one-half wavelength apart and fed with equal, syn-phased currents as in fig. 9a. The polar diagram in the equatorial plane may be obtained in the following manner. Divide the array into two pairs of aerials; the polar diagram of each of these pairs is, by the previous paragraph, the elongated figure-of-eight shown in diagram A 5 of fig. 8, and repeated in fig. 9b. For the four aerials spaced $\frac{\lambda}{2}$ apart, we may now substitute two aerials, each having the above polar diagram, but spaced one wavelength apart as in fig. 9c. According to diagram A 9 of fig. 8, two dipoles with this spacing, and syn-phased currents, have a polar diagram with four lobes; this diagram is repeated in fig. 9d.

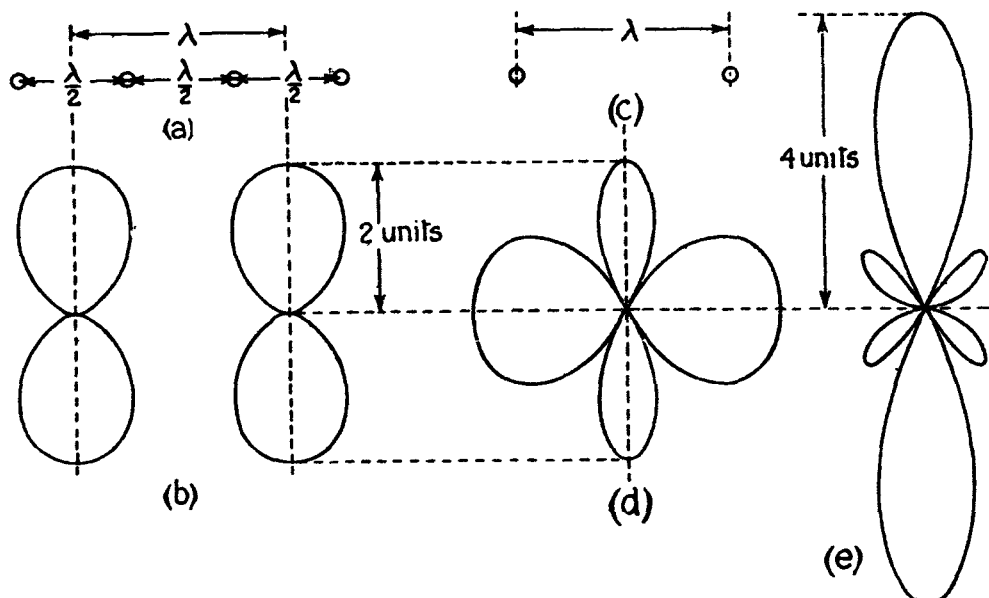


FIG. 9, CHAP. XV.—Development of polar diagram for four parallel dipoles carrying syn-phased current.

The field distribution of the combination of two radiators which, individually, give the figure-of-eight diagram, fig. 9b, may now be obtained by multiplying together the corresponding polar radii of figs. 9b and 9d, giving the result shown in fig. 9e. The principle of combining parallel aerials carrying syn-phased currents is the basis of what are called broadside arrays.

19. The above process may obviously be extended to obtain the field distribution in the equatorial plane for any number of dipoles irrespective of the spacing and the phase of current in the respective aerials. Thus, suppose we have an array of four parallel dipoles A B C D, fig. 10a, spaced $\frac{\lambda}{4}$ apart, each carrying a current of I amperes. Let the current in B lead by 90° on that in A, the current in C lead by 90° on that in B and so on. The polar diagram for the pair A and B, or for the pair C and D, is given in fig. 8, diagram C 3, which is reproduced in fig. 10b; it

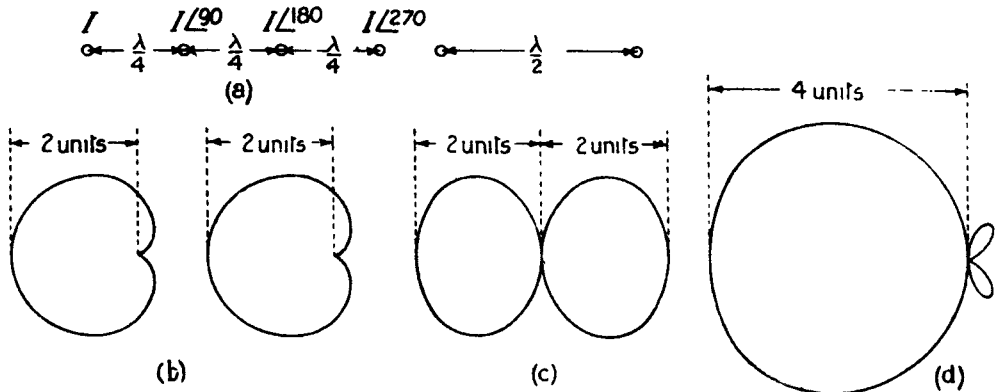


FIG. 10, CHAP. XV.—Development of polar diagram for four parallel dipoles with currents in progressive phase difference of 90° .

is a cardioid or heart-shape. For the four dipoles, we may now substitute two radiators each having this polar diagram, spaced $\frac{\lambda}{2}$ apart, and with currents in anti-phase. Two dipoles with this spacing and current phase give the polar diagram E 5, fig. 8, reproduced in fig. 10c. The polar diagram of the four-element array is found by multiplying together the corresponding polar radii of diagrams C 3 and E 5, resulting in the diagram shown in fig. 10d. It will be observed that the array is substantially uni-directional, maximum radiation being directed in the direction of the aerial in which the phase of the current is lagging. The principle of combining parallel aerials carrying currents differing by a constant angle, which in turn is related to the spacing of the elements, is the basis of what are called end-fire aerial arrays.

Radiation in the plane of the aerials.

20. We may now investigate the shape of the polar diagram in the plane containing the aerials. In fig. 11 let A and B be two parallel dipoles each carrying a loop current $i = \mathcal{I}_0 \cos \omega t$. At a point P at an angle φ above the equatorial plane, situated at a distance r from the origin O, where $r \gg \lambda$, the fields will be

$$\begin{aligned} r_A &= K \mathcal{I}_0 \frac{\cos \left(\frac{\pi}{2} \sin \varphi \right)}{\cos \varphi} \cos \left[\omega t + \frac{\pi}{2} - \frac{2\pi}{\lambda} \left(r + \frac{d}{2} \cos \varphi \right) \right] \\ r_B &= K \mathcal{I}_0 \frac{\cos \left(\frac{\pi}{2} \sin \varphi \right)}{\cos \varphi} \cos \left[\omega t + \frac{\pi}{2} - \frac{2\pi}{\lambda} \left(r - \frac{d}{2} \cos \varphi \right) \right] \end{aligned}$$

CHAPTER XV.—PARA. 20

The Current Distribution Factor $\frac{\cos \left(\frac{\pi}{2} \sin \varphi \right)}{\cos \varphi}$

has previously been introduced to account for the fact that the dipole does not radiate uniformly in any plane perpendicular to the equatorial. The combined field is $\gamma_{\varphi} = \gamma_A + \gamma_B$ and

$$\gamma_{\varphi} = \left[2K \mathcal{J}_0 \frac{\cos \left(\frac{\pi}{2} \sin \varphi \right)}{\cos \varphi} \cos \left(\frac{\pi d}{\lambda} \cos \varphi \right) \right] \cos \left(\omega t + \frac{\pi}{2} - \frac{2\pi}{\lambda} r \right). \quad (13)$$

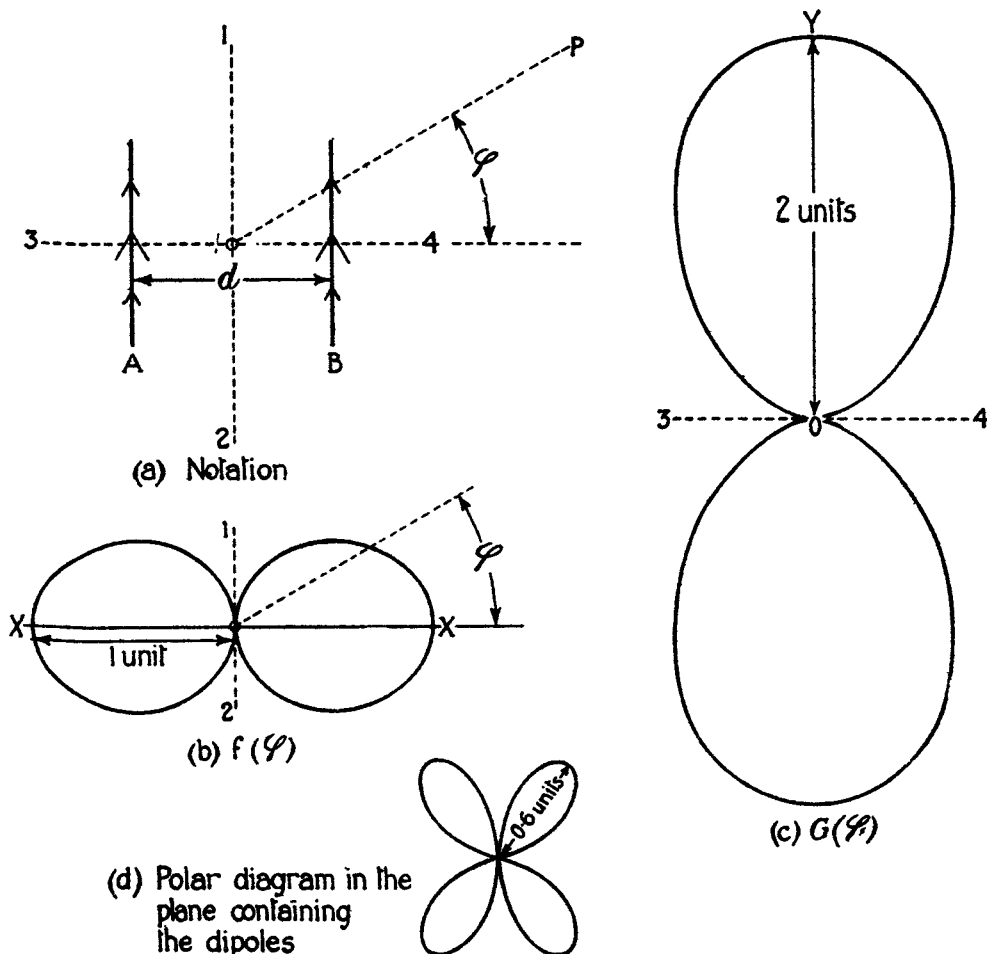


FIG. 11, CHAP. XV.—Radiation in plane containing two parallel syn-phased dipoles.

The portion of the right-hand member which is enclosed in square brackets is the amplitude of the field in the direction O P. It is the product of three factors

$$(i) K \mathcal{J}_0 = \frac{60}{r} \mathcal{J}_0.$$

$$(ii) \frac{\cos \left(\frac{\pi}{2} \sin \varphi \right)}{\cos \varphi}, \text{ i.e. the Current Distribution Factor of a } \frac{\lambda}{2} \text{ dipole.}$$

$$(iii) 2 \cos \left(\frac{\pi d}{\lambda} \cos \varphi \right), \text{ i.e. the Grating Factor for a pair of syn-phased dipoles in the plane containing them. This is of exactly the same form as the equatorial plane Grating Factor, but is a function of the angle } \varphi \text{ instead of being a function of the angle } \theta.$$

The Current Distribution Factor is given by fig. 6, and the Grating Factor by the upper row of diagrams in fig. 8. Thus the resultant amplitude in any particular case may be obtained by multiplying the constant $\frac{60}{r} \mathcal{J}_0$ by two polar radii obtainable from the diagrams. As an example,

take $d = \frac{\lambda}{2}$. Fig. 11b is the Current Distribution Factor, fig. 11c the Grating Factor, and the diagram obtained from the polar products is shown in fig. 11d. This product has a maximum value of 0.6, at an angle of approximately 55° . If the Grating Factor in this plane is denoted by $G(\varphi)$, the R.M.S. field in the plane containing the aerials is

$$E_\varphi = \frac{60}{r} I_0 \times F \times f(\varphi) \times G(\varphi)$$

where F and $f(\varphi)$ are the Form and Current Distribution Factors as before. It will be seen later that if the co-ordinates of the point P are r, θ, φ , the Grating Factor becomes

$$G(\theta, \varphi) = 2 \cos \left(\frac{\pi d}{\lambda} \cos \theta \cos \varphi \right).$$

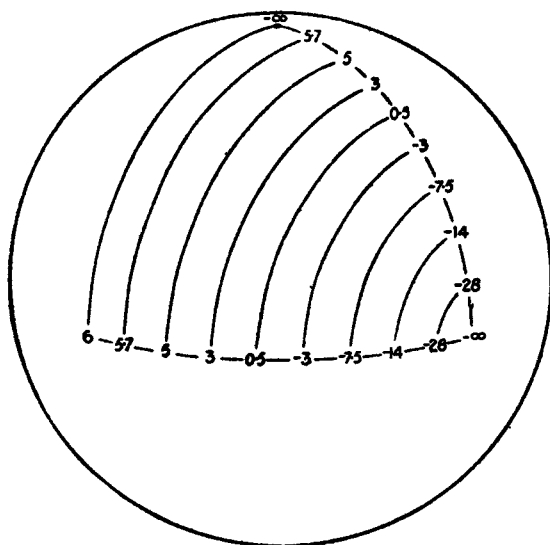
Three dimensional polar diagram

21. We have now shown how to obtain the polar diagrams for a pair of spaced aerials in the equatorial plane and in the perpendicular plane containing the aerials. While it is possible to calculate the polar diagram of any combination of aerials in all directions in space, the process becomes very tedious when more than two or three aerials are involved. The solid polar diagram may however be obtained by combining the diagrams for the equatorial plane and that containing the aerials. The process will be illustrated by taking the two parallel dipoles, spaced $\frac{\lambda}{2}$ apart in free space as before (fig. 11a) and carrying equal, syn-phased currents. The R.M.S. field strength at an angle φ with respect to the equatorial plane is proportional to the Current

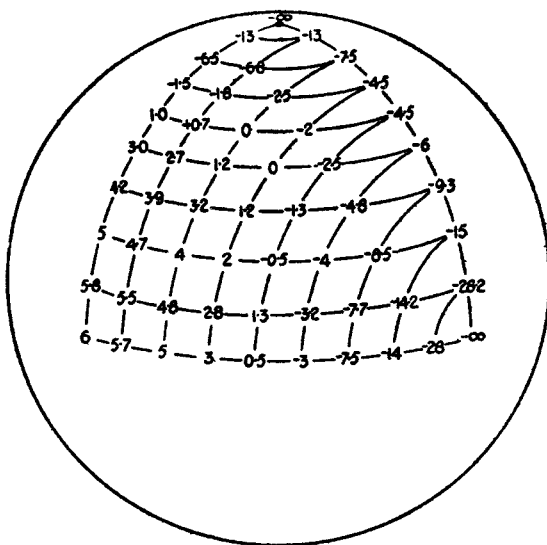
Distribution Factor $f(\varphi)$; in this particular case $f(\varphi) = \frac{\cos \left(\frac{\pi}{2} \sin \varphi \right)}{\cos \varphi}$ and has already been plotted in fig. 6. If this diagram is rotated about the axis 1,2, fig. 11a, the resulting solid figure (when multiplied by $\frac{60}{r} I_0$) gives the three-dimensional diagram of a single dipole. The combination of two such dipoles introduces a Grating Factor which has been shown to be $2 \cos \left(\frac{\pi d}{\lambda} \cos \theta \right)$ in the equatorial plane and $2 \cos \left(\frac{\pi d}{\lambda} \cos \varphi \right)$ in the plane containing the aerials.

CHAPTER XV.—PARA. 21

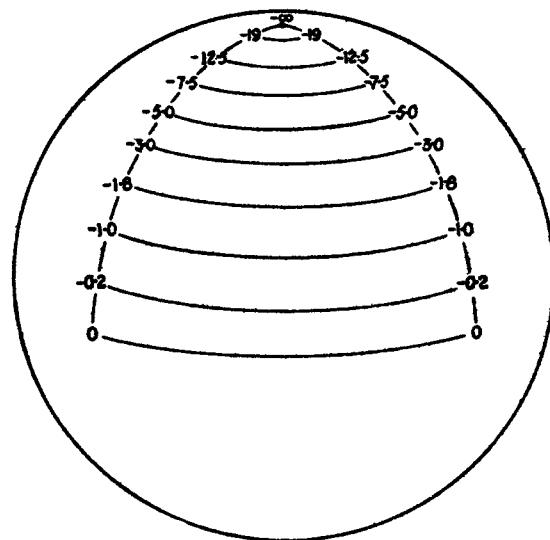
The Grating Factor is plotted in fig. 11c ; if it is rotated about the axis 3, 4, the result is another solid figure. The polar radii of the latter, for any direction in space, gives a factor by which the quantity $\frac{60}{r} I_0 f(\varphi)$ must be multiplied, in order to give the R.M.S. field at any particular point. The radius O X in fig. 11b is equal to unity, and the radius O Y in fig. 11c is equal to two units. Thus, in the equatorial plane, the field strength in a direction perpendicular to the line upon which the dipoles are situated is twice that of a single dipole.



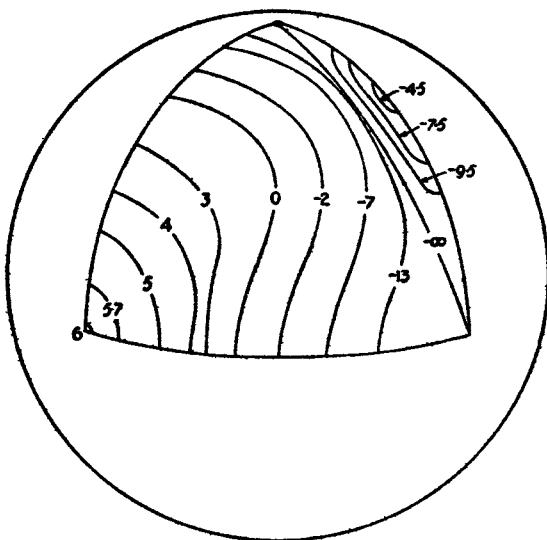
(a) Sphere with parallels corresponding to grating factor $G(\varphi)$ in db



(c) Sum of $G(\varphi)$ and $f(\varphi)$ in db



(b) Sphere with parallels corresponding to C.D. factor $f(\varphi)$ in db



(d) Contour lines derived from (c)

FIG. 12, CHAP. XV.—Gain contours on spherical surface.

22. Instead of in units of length, the radii may be expressed in decibels above or below unity. The field in any direction is then given, in decibels above or below the equatorial field of a single dipole, merely by adding the decibels corresponding to the respective radii of figs. 11b and 11c. If we take a spherical surface and mark off a number of equal zones parallel to the equatorial plane, each of the boundary lines between adjacent zones may be marked to show the number of decibels below the field strength in the equatorial plane as in fig. 12b, the figures being derived from the Current Distribution Factor. Similarly, if we draw a number of equal zones in a plane perpendicular to the equatorial plane and to the plane containing the aerials, the boundary lines between these zones may also be marked in decibels above or below unity as in fig. 12a, the figures being obtained from the Grating Factor. At the points of intersection of any two lines, the field strength is above or below the equatorial field of a single aerial by the sum of the decibels appropriate to the two intersecting lines. One quadrant of a spherical surface, with both sets of zones superimposed, is shown in fig. 12c. It must be particularly noted that although the boundary lines in fig. 12b correspond to parallels of latitude, the boundary lines of fig. 12a do not pass through the pole and are not analogous to meridians of longitude.

23. If now we insert, at the intersection of all boundary lines in the spherical surface, a number equal to the algebraic sum of the decibels appertaining to the two intersecting lines, we obtain the gain or loss in decibels compared with the standard at different points on the sphere. These figures have been inserted in fig. 12c. By joining all points of equal gain, we obtain a field strength contour diagram as in fig. 12d. A close examination of this figure shows that in each quadrant of the surface there are two maxima. One, corresponding to the main lobe, is 6 db. above the standard while the other is at an elevation of about 55° , in the plane containing the aerials, and its maximum is about 4 db. below the standard. This lobe has already been found to exist (paragraph 20). From these data we may make a solid model of the polar diagram in plasticine, as shown in fig. 13. To do this, the gain in db. above or below the standard must be converted back to absolute field strength.

Field strength map—the sinusoidal projection

24. When the principles involved in the production of the solid diagram are thoroughly appreciated, it will be found easy to construct a map showing the gain or loss in different directions. We may consider the aerial array to be situated at the centre of a sphere and to illuminate

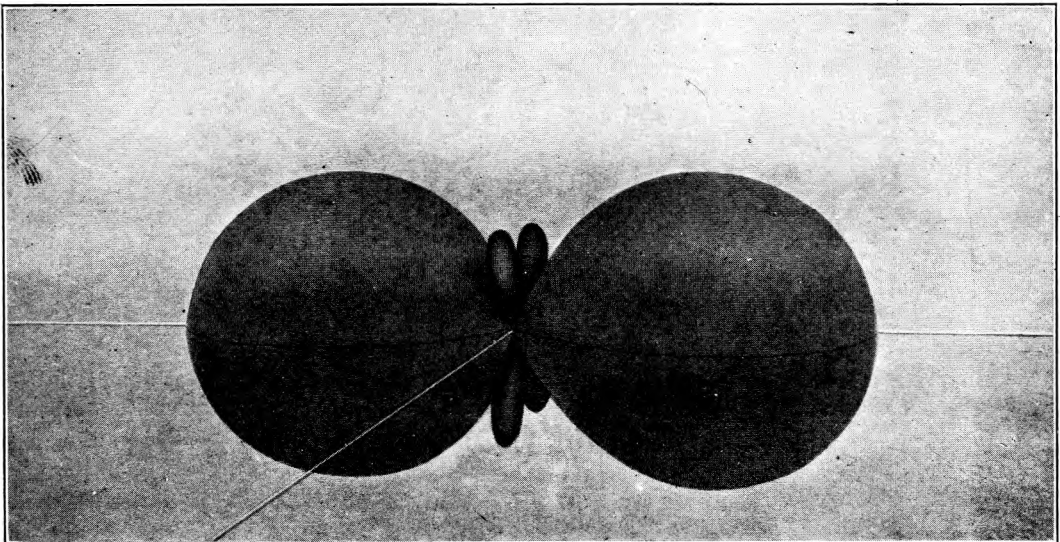


FIG. 13, CHAP. XV.—Solid polar diagram—parallel dipoles in free space.

CHAPTER XV.—PARA. 25

different regions with greater or less intensity. The delineation of a spherical surface upon a plane is most familiar in the form of Mercator's projection of the earth. This projection is unsuitable for general use in aerial array theory because the high latitudes cannot be shown with accuracy, and it may be necessary to show the field strength vertically over the aerial. A suitable projection is that known as sinusoidal, in which the length of a parallel of latitude is proportional to the cosine of the latitude. This is shown in fig. 14. In the original drawing, the length of one quadrant of a parallel of latitude in the equatorial plane, i.e. latitude 0° , is 9 inches. The length of the corresponding quadrant in latitude 10° is $9 \cos 10^\circ = 8.85$ inches, in latitude 20° is 8.45 inches and so on. The meridian corresponding to longitude 0° (with the convention of fig. 10) is a line through the points given above and is a cosine curve. Longitudes 10° , 20° , etc., are also cosine curves obtained by the division of each 90° into nine equal parts. Once the sinusoidal graticule has been prepared, the

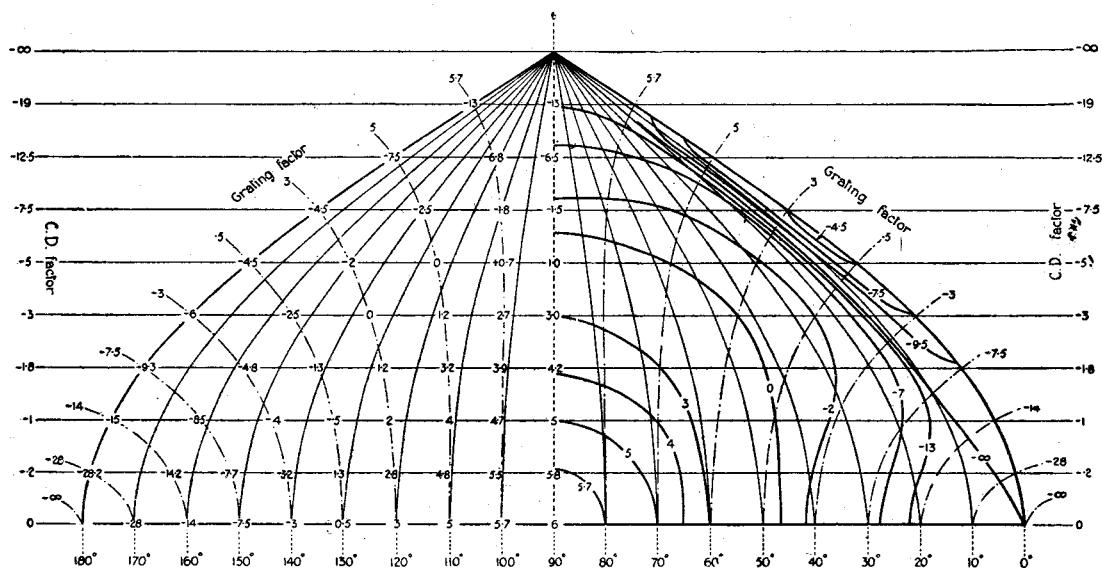


FIG. 14, CHAP. XV.—Gain contours shown on sinusoidal projection.

parallels of latitude may be allotted the appropriate Current Distribution Factors (in decibels) and the zones perpendicular to these may be inserted by freehand drawing with sufficient accuracy for most purposes. In fig. 14 these are denoted by chain-dotted lines. The latter are allotted their appropriate values of Grating Factor (in decibels). The total gain is then inserted at the intersecting points, and the gain contours drawn. Alternatively, the gain in decibels may be reconverted to absolute field strength. There is of course no objection to working in absolute field strength from the beginning, but this would necessitate finding the product of two numbers for each intersecting point.

Use of vector algebra

25. When it is necessary to calculate the field at a point P in space, having the co-ordinates r, θ, ϕ , the method now to be described will be found more convenient for algebraic purposes than the purely trigonometric methods previously adopted. Instead of considering the sinusoidal current as the product of a constant, i.e. the amplitude $\mathcal{I}$, and a trigonometrical function of time, e.g. $\cos \omega t$, it is considered as the product of a vector $\mathbf{I}$ and a vector operator $\epsilon^{j\omega t}$. Now $\epsilon^{j\omega t} = \cos \omega t + j \sin \omega t$ so that $\mathbf{I}\epsilon^{j\omega t} = \mathbf{I} \cos \omega t + j \mathbf{I} \sin \omega t$. Since in operations involving complex quantities of this kind, the real and imaginary parts are entirely independent, we may

deal with a current $i = \mathcal{I} \cos \omega t$ by saying "let i be the real part of $\mathbf{I} e^{j\omega t}$." The magnitude of the vector $\mathbf{I}$ is of course equal to the amplitude $\mathcal{I}$ of the current, in fact $\mathbf{I}$ may be regarded as the product of the scalar $\mathcal{I}$ and a unit vector. In practice, it is usual to write merely "let $i = \mathbf{I} e^{j\omega t}$ " the real part only of the final result being taken. For example, consider the field due to a $\frac{\lambda}{2}$ dipole in free space, carrying a loop current $i_0 = \mathcal{I}_0 \cos \omega t$. Let this current be

$\mathbf{I}_0 e^{j\omega t}$ Then the field γ_φ at the point $P = r, \varphi$, will be

$$\gamma_\varphi = \frac{60}{r} f(\varphi) \mathcal{I}_0 \cos \left(\omega t + \frac{\pi}{2} + \frac{2\pi}{\lambda} r \right) \quad \dots \dots \dots (14a)$$

in the notation of previous paragraphs. In the present notation,

$$\gamma_\varphi = \frac{60}{r} f(\varphi) \mathbf{I}_0 e^{j\omega t} e^{j\frac{\pi}{2}} e^{j\frac{2\pi}{\lambda} r} \quad \dots \dots \dots (14b)$$

or more economically

$$\gamma_\varphi = \frac{60}{r} f(\varphi) \mathbf{I}_0 e^{j \left(\omega t + \frac{\pi}{2} + \frac{2\pi}{\lambda} r \right)} \quad \dots \dots \dots (14c)$$

26. The advantage of this notation lies in the ease with which the fields due to two or more radiators can be combined, even if they differ both in magnitude and phase. Referring to Chapter V, an impedance of magnitude $Z = \sqrt{R^2 + X^2}$ ohms may be represented both in magnitude and in its effect on the phase angle of a current, in any of the following ways,

$$Z \angle \theta = R + jX = Z e^{j\theta}$$

or

$$Z \angle \theta = R - jX = Z e^{-j\theta}$$

$$\left(\theta = \tan^{-1} \frac{X}{R} \right)$$

depending upon whether the reactance is positive (inductive) or negative (capacitive). Obviously $Z \angle \theta = Z \angle -\theta$ and vice versa. For example, an E.M.F. $\mathbf{E} e^{j\omega t}$ acting in a circuit of $Z \angle \theta$ ohms, will produce an instantaneous current,

$$i = \frac{\mathbf{E} e^{j\omega t}}{Z \angle \theta} = \frac{\mathbf{E} e^{j\omega t}}{Z} \angle -\theta$$

$$= \frac{\mathbf{E}}{Z} e^{j\omega t} e^{-j\theta}$$

$$= \frac{\mathbf{E}}{Z} e^{j(\omega t - \theta)} \text{ (real part only)}$$

$$= \frac{\mathcal{E}}{Z} \cos(\omega t - \theta),$$

CHAPTER XV.—PARA. 27

i.e. a positive reactance produces a lagging current. Similarly, an impedance $Z \sqrt{\theta}$ ohms, acting under the same conditions, produces an instantaneous current

$$\begin{aligned} i &= \frac{E e^{j\omega t}}{Z \sqrt{\theta}} = \frac{E e^{j\omega t}}{Z} \frac{1}{\sqrt{\theta}} \\ &= \frac{E}{Z} e^{j(\omega t + \theta)} \text{ (real part only)} \\ &= \frac{E}{Z} \cos (\omega t + \theta), \end{aligned}$$

i.e. a negative reactance produces a leading current. It is also convenient to adopt a distinctive type of symbol for any vector operator, for use where it is unnecessary or impossible to define its properties completely. In the following text such operators will be denoted by lower-case (i.e. "small") Clarendon type, thus naturally associating with vector quantities. The latter are printed in Clarendon capitals, except where the symbol is a Greek character, when a bar superior is used thus $\bar{T}$. Vector operators are often used to denote the vector ratio between two quantities as follows:—Suppose we have two currents, $i_1 = \mathcal{I}_1 \cos (\omega t - \theta)$ and $i_2 = \mathcal{I}_2 \cos (\omega t + \varphi)$. The ratio $\frac{\mathcal{I}_2}{\mathcal{I}_1}$ of the amplitudes is a mere number and may be denoted by M . We also require to know their relative phase, and in the vector notation

$$\begin{aligned} i_1 &= I_1 e^{j(\omega t - \theta)} = I_1 / \omega t - \theta \\ i_2 &= I_2 e^{j(\omega t + \varphi)} = I_2 / \omega t + \varphi \\ \frac{i_2}{i_1} &= \frac{I_2 / \omega t + \varphi}{I_1 / \omega t - \theta} \\ &= \frac{I_2}{I_1} \frac{\omega t + \varphi - \omega t + \theta}{\omega t - \theta} \\ &= M / \varphi - \theta' \\ &= m. \end{aligned}$$

The above method of treatment leads to the simple algebraic solution of problems which would otherwise be comparatively difficult and much more tedious.

General case of two parallel radiators

27. Referring to fig. 15, let A and B be two parallel but not necessarily identical aerials. Their midpoints are, however, equally spaced on either side of an origin O in the equatorial phase, the distance apart being d . Consider the field at the point P = r, φ, θ , where $r = OP$ is very much greater than d . Let the angle XOP = α . Then XAP = XBP = α also. We may therefore write

$$\begin{aligned} AP &= r_A = r + \frac{d}{2} \cos \alpha \\ BP &= r_B = r - \frac{d}{2} \cos \alpha \end{aligned}$$

with negligible error. Now suppose the currents at the midpoints to be $I'_A = I_A e^{j\omega t}$ and $I'_B = I_B e^{j(\omega t + \beta)}$. If F_A, F_B are the Form Factors and $f_A(\varphi), f_B(\varphi)$, the Current Distribution Factors of the respective aerials with regard to their midpoints, the individual fields due to the two aerials will be

$$\gamma_A = j \frac{60}{r} F_A \cdot f_A(\varphi) I'_A e^{-j \frac{2\pi}{\lambda} r_A}$$

$$\gamma_B = j \frac{60}{r} F_B \cdot f_B(\varphi) I'_B e^{-j \frac{2\pi}{\lambda} r_B}$$

or

$$\gamma_A = j \frac{60}{r} F_A \cdot f_A(\varphi) I'_A e^{-j \frac{2\pi}{\lambda} r} e^{-j \frac{\pi d}{\lambda} \cos \alpha}$$

$$\gamma_B = j \frac{60}{r} F_B f_B(\varphi) I'_B e^{-j \frac{2\pi}{\lambda} r} e^{+j \frac{\pi d}{\lambda} \cos \alpha}$$

Thus the combined field is

$$\gamma = j \frac{60}{r} e^{-j \frac{2\pi}{\lambda} r} \left[F_A \cdot f_A(\varphi) I'_A e^{-j \frac{\pi d}{\lambda} \cos \alpha} + F_B \cdot f_B(\varphi) I'_B e^{+j \frac{\pi d}{\lambda} \cos \alpha} \right] \dots \dots \dots (15)$$

The above process may obviously be applied to an array consisting of any number of elements.

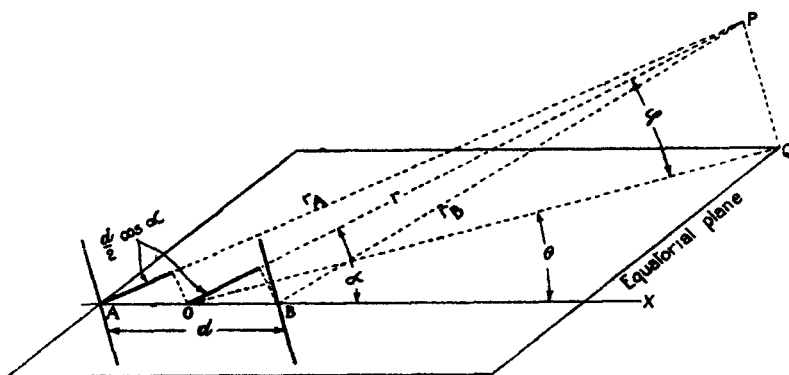


FIG. 15, CHAP. XV.—Notation, general case of parallel radiators.

28. Let us now take the specific case where the aerials are identical. Then $F_A \cdot f_A(\varphi) = F_B \cdot f_B(\varphi) = F \cdot f(\varphi)$.

For brevity we may write $K = j \frac{60}{r} F \cdot f(\varphi) e^{-j \frac{2\pi}{\lambda} r}$, and proceed to allow for the difference in

magnitude and phase of the currents. Let the amplitudes be $\mathcal{I}_A, \mathcal{I}_B$, and $\mathcal{I}_B = M \mathcal{I}_A$. The current in the aerial B leads on the current in the aerial A by an angle β , or $I'_B = M \angle \beta I'_A$.

CHAPTER XV.—PARAS. 29-30

The total field is therefore

$$\begin{aligned} \gamma &= K \left[I'_A \varepsilon^{-j \frac{\pi d}{\lambda} \cos \alpha} + M / \beta I'_A \varepsilon^{+j \frac{\pi d}{\lambda} \cos \alpha} \right] \\ &= K I'_A \left[\varepsilon^{-j \frac{\pi d}{\lambda} \cos \alpha} + M \varepsilon^{+j \left(\beta + \frac{\pi d}{\lambda} \cos \alpha \right)} \right] \\ &= K I'_A \varepsilon^{-j \frac{\pi d}{\lambda} \cos \alpha} \left[1 + M \varepsilon^{+j \left(\beta + \frac{2\pi d}{\lambda} \cos \alpha \right)} \right]. \quad \dots \quad (16) \end{aligned}$$

Now $K I'_A \varepsilon^{-j \frac{\pi d}{\lambda} \cos \alpha}$ is the field due to the aerial A alone and may be denoted by γ_A . Then

$$\begin{aligned} \gamma &= \gamma_A \left[1 + M \varepsilon^{+j \left(\beta + \frac{2\pi d}{\lambda} \cos \alpha \right)} \right] \\ &= \gamma_A \left[1 + M / \beta + \frac{2\pi d}{\lambda} \cos \alpha \right]. \end{aligned}$$

It now only remains to find the scalar value of the quantity enclosed in brackets. To do this let

$$\beta + \frac{2\pi d}{\lambda} \cos \alpha = \psi.$$

$$M / \psi = M \cos \psi + j M \sin \psi.$$

The required scalar is that of $1 + M / \psi$ or $1 + M \cos \psi + j M \sin \psi$, and from Chapter V this is known to be $\sqrt{(1 + M \cos \psi)^2 + M^2 \sin^2 \psi}$. This easily reduces to $\sqrt{1 + 2M \cos \psi + M^2}$ and therefore

$$\gamma = \gamma_A \sqrt{1 + 2M \cos \psi + M^2}. \quad \dots \quad (17)$$

The R.M.S. field will be

$$F = \frac{60}{r} F.f(\varphi) I_A \sqrt{1 + 2M \cos \psi + M^2}. \quad \dots \quad (18)$$

29. Before proceeding further, let us examine the angle α , which is more conveniently expressed in terms of the angles θ , φ . An examination of fig. 15 shows that the projection O Q of O P upon the equatorial plane is O P $\cos \varphi$, and that the projection of O Q upon the datum line O X is O Q $\cos \theta$. Since O P = r , O Q = $r \cos \varphi$ and the projection of O P upon O X is $r \cos \varphi \cos \theta$. The direct projection of O P upon O X is obviously O P $\cos \alpha$ or $r \cos \alpha$, i.e.

$$r \cos \alpha = r \cos \varphi \cos \theta$$

$$\cos \alpha = \cos \varphi \cos \theta.$$

30. We are now in a position to discuss the polar diagram. With unequal currents in the aerials, the only satisfactory basis of comparison with a single aerial is for equal power. From paragraph 14 we know that the powers radiated by the aerials A and B will be $I_A^2 R_A$ and $I_B^2 R_B$ respectively, where R_A and R_B are the radiation resistances. With identical aerials $R_A = R_B$ and the total radiated power is

$$P_T = I_A^2 R_A + (M I_A)^2 R_A. \quad \dots \quad (19)$$

For a given power, then, the current I_A must be equal to $\sqrt{\frac{P_T}{R_A(1 + M^2)}}$.

Inserting this value for the current, we have, for the R.M.S. field

$$F = \frac{60}{r} F \cdot f(\varphi) \sqrt{\frac{P_T}{R_A (1 + M^2)}} \sqrt{1 + M^2 + 2M \cos \left(\beta + \frac{2\pi d}{\lambda} \cos \varphi \cos \theta \right)} \quad \dots \dots \dots (20)$$

The polar diagram of a single aerial of the same kind, situated at the origin O, is given by the expression

$$F = \frac{60}{r} F \cdot f(\varphi) \sqrt{\frac{P_T}{R_A}} \quad \dots \dots \dots (21)$$

and is a circle about the origin. The ratio of the field produced by the two aerials, to that produced by a single aerial radiating an equal power, is

$$G(\varphi, \theta) = \sqrt{\frac{1 + M^2 + 2M \cos \left(\beta + \frac{2\pi d}{\lambda} \cos \varphi \cos \theta \right)}{1 + M^2}}$$

31. From this expression it is possible to calculate diagrams similar to those of fig. 8, for any value of M/β and at any angle φ with respect to the equatorial plane. It is obviously impossible to portray all the possibilities here; equatorial plane diagrams corresponding with fig. 8 are derived by putting $M = 1$, $\varphi = 0$, and letting β take any required value. With these substitutions

$$\begin{aligned} G(\theta) &= \frac{\sqrt{2 + 2 \cos \left(\beta + \frac{2\pi d}{\lambda} \cos \theta \right)}}{\sqrt{2}} \\ &= \sqrt{1 + \cos \psi}. \end{aligned}$$

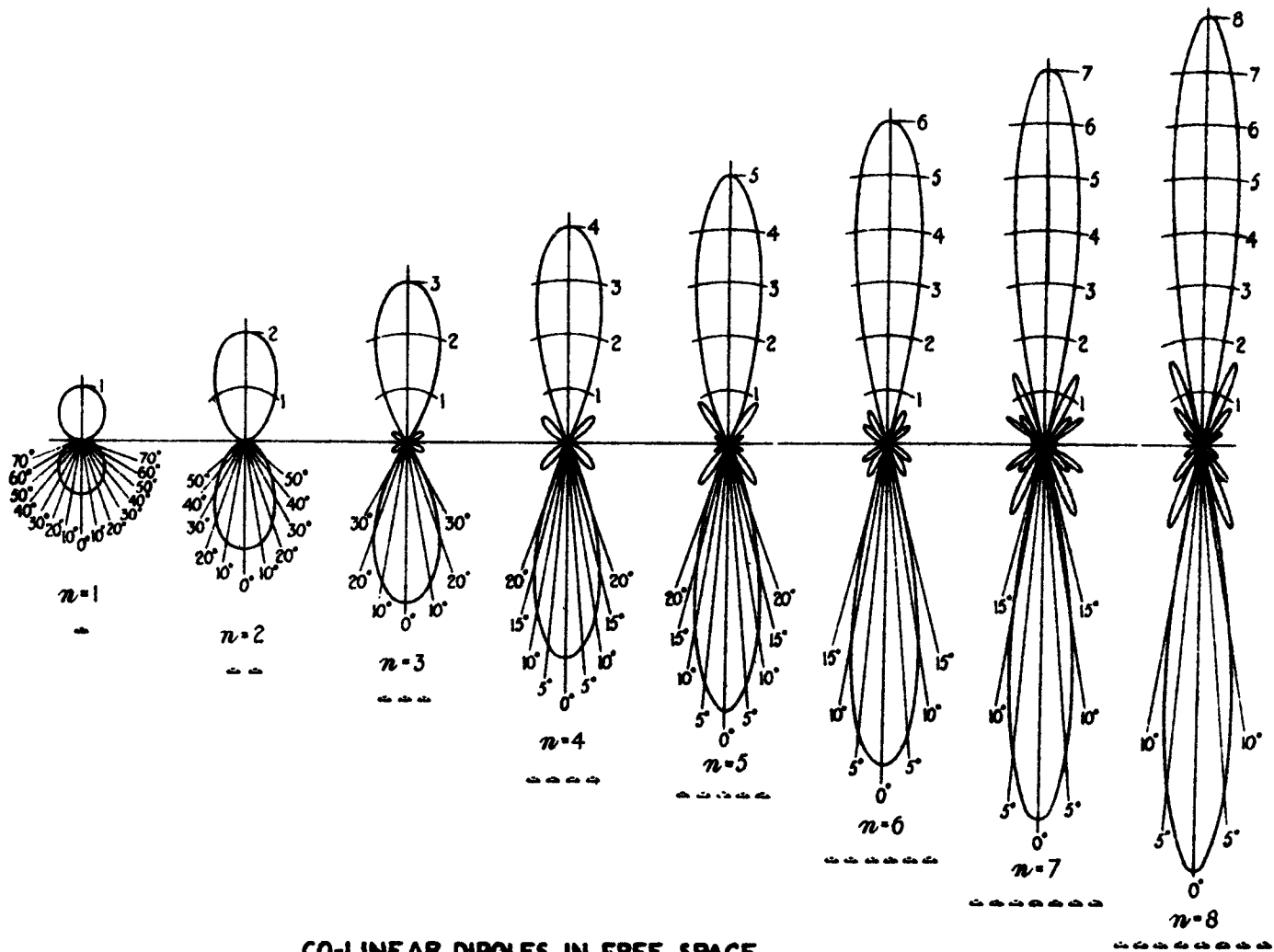
By trigonometry $1 + \cos \psi = 2 \cos^2 \frac{\psi}{2}$ so that

$$\begin{aligned} G(\theta) &= \sqrt{2} \cos \frac{\psi}{2} \\ &= \sqrt{2} \cos \left(\frac{\beta}{2} + \frac{\pi d}{\lambda} \cos \theta \right). \end{aligned}$$

This expression is the same as that developed by a different method in paragraph 17 except that the factor $\sqrt{2}$ appears instead of 2. This is because we have obtained the Grating Factor for equal power in single aerial and array respectively, whereas in paragraph 17 the power in the array was four times that in the single aerial.

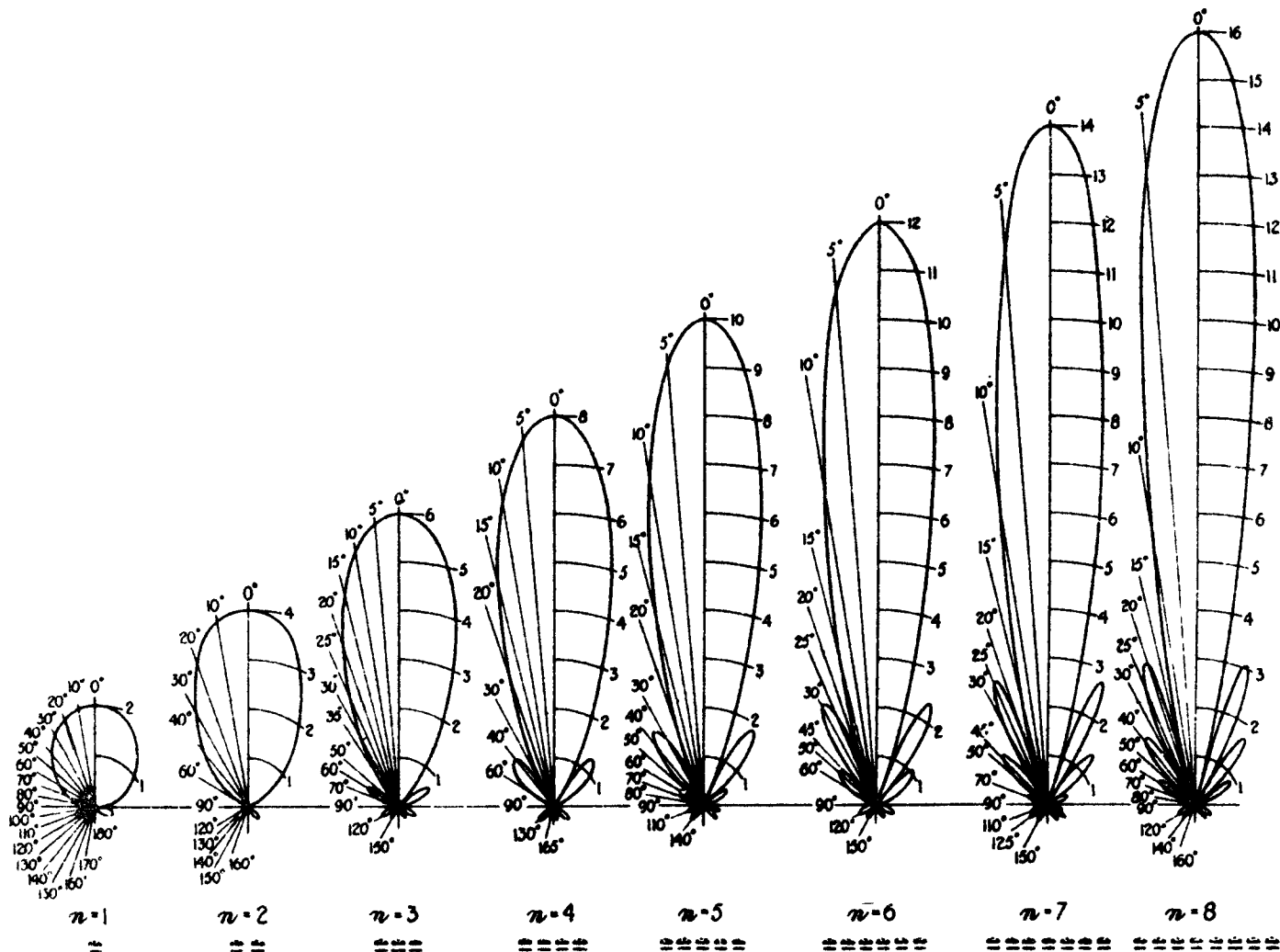
Co-linear dipoles

32. Instead of being placed parallel to each other, single wire radiators, particularly $\frac{\lambda}{2}$ dipoles, are sometimes placed end to end as shown in fig. 16, and are then said to be co-linear. The polar diagram of a simple array consisting of two co-linear dipoles can be found as follows. In fig. 16 let all measurements be made from an origin O lying between the two dipoles and on their common axis, and let their current loops be separated by a distance d . For simplicity let



CO-LINEAR DIPOLES IN FREE SPACE

FIG.17
CHAP.XV.



CO-LINEAR DIPOLES WITH REFLECTORS

can be found by the methods explained with reference to parallel dipoles. Fig. 17 shows the product $f(\varphi) G(\varphi)$ for all numbers up to 8 co-linear dipoles. This figure has an important bearing on the radiation from arrays of horizontal dipoles.

Reflector aerial

33. Hitherto, in considering the radiation from two parallel aerials, we have ignored the effect of one aerial upon the other. It is obvious that when both are supplied with energy each will receive energy from the other. Of the received energy, a portion is converted into heat and the remainder radiated into space. Now let us consider two parallel dipoles, A and B, $\frac{\lambda}{4}$ apart in free space, and consider their radiation in the equatorial plane, when A is supplied with a current $i = \mathcal{I}_A \cos \omega t$, and B is unenergized. Then the field due to the aerial A will induce an E.M.F. in B, and consequently an oscillatory current of the same frequency. This "induction" is really caused by both the induction and the radiation fields of A, but for the present we shall neglect the former. The magnitude $\mathcal{I}_B$ of the induced current in B, under these conditions, will be equal to that of the current in A, but $\mathcal{I}_B$ will lead on $\mathcal{I}_A$ by 90° . The aerials A and B therefore radiate an equal amount of energy per second, and the field at any point can be calculated as in previous paragraphs, putting the angle β equal to 90° ; the polar diagram in the equatorial plane is given by fig. 8, C 3. The effect in the plane containing the aerials is somewhat similar, the only modification being due to the Current Distribution Factor. For comparison with fig. 17, fig. 18 gives the polar diagram in this plane of co-linear dipoles with reflectors, calculated on the above assumptions.

34. The exact manner in which a reflector aerial will function depends upon three factors; first, its distance from the energized aerial; second, the ratio of its induced current to that in the energized aerial; third, whether it is reactive or non-reactive, i.e. tuned or untuned to the frequency of the energized aerial. The second factor is obviously not independent of the third. Reference to fig. 8 shows that if the currents in the energized and reflector aerials are equal, the polar diagram is more or less uni-directional whenever the phase difference between the two currents is greater than 0° and less than 180° , provided that the spacing is less than $\frac{\lambda}{2}$. Particular attention is directed to the diagrams B 2, B 3, B 4, C 2, C 3, C 4, D 2, D 3, D 4, of fig. 8, which show the theoretical possibilities which may arise. The effect of the induction field cannot be entirely ignored, and will receive further consideration.

Effect of the ground

35. In practice, the field produced at a given point by a transmitting aerial is always affected by the presence of the earth's surface, but a complete treatment allowing for the curvature of the earth, and its finite conductivity and permittivity, is extremely complex. For many purposes, however, the earth may be considered as a flat, perfectly conductive surface, and this simplification is of great help in visualizing the nature of the effect. A perfectly conducting earth would act as a perfect reflector of electro-magnetic waves, and a flat earth as a plane reflector. If then we consider the earth's surface in the immediate vicinity of the aerial to be both perfectly conductive and plane, we may treat certain problems by a method analogous to that used in geometrical optics, i.e. by supposing the reflector to give a virtual image of the actual radiator. The virtual image is defined as a point from which rays appear to diverge after reflection, although no rays actually pass through the point. This conception is illustrated in fig. 19a which shows a hertzian doublet A B situated above a perfectly reflecting earth. At the instant depicted, the current is flowing from A to B, and consequently a positive charge is accumulating at B.

36. Now take a point B' situated at a distance $OB' = OB$ on the other side of the reflecting surface, so that B O B' is straight and perpendicular to the surface. B' is then the geometrical virtual image of B. Similarly, we may locate the geometrical virtual image A' of the point A. Now considering B to be a small sphere, it must possess capacitance with respect to the perfectly conducting surface and if it carries a positive charge, its electrostatic field would be distributed somewhat as shown in fig. 19b. The field due to a similar charge of opposite polarity, situated at the point B', is also shown, and it is seen, in conjunction with fig. 19a, that a positive charge

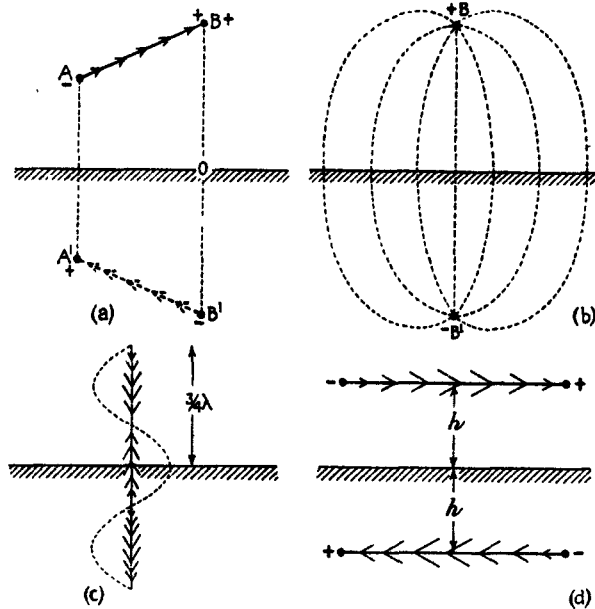


FIG. 19, CHAP. XV.—Illustrations of image theorem.

at B, above a perfect reflector, will have the same field as would be set up if the reflecting surface were removed and an equal and opposite charge placed at the point B'. Similar considerations apply to the points A and A' and therefore, when a current is flowing from A to B, an equal current must be considered to flow from B' to A'. The field strength at any point above the reflector is found by algebraic addition of the fields due to the aerial itself and to its virtual image, with due regard to the direction of current in the latter, as determined by the above considerations.

Fig. 19c shows the distribution of current in a vertical $\frac{3}{4}\lambda$ aerial, and the current in a horizontal dipole is shown in fig. 19d together with those of the images. In certain circumstances the image theorem lends itself to comparatively simple application, and it will now be applied to find the radiation characteristics of the horizontal dipole.

Horizontal dipole

37. For purposes of notation; the dipole is shown in figs. 20a and 20b. We shall consider the vertical polar diagram in the plane of the latter figure. By the preceding paragraph the image A' B' of the dipole A B is as shown, and therefore we require to find the polar diagram of two parallel dipoles, $d = 2h$ apart, carrying currents in phase opposition. If γ_A is the field due to the aerial at the point P, and γ_R the field due to the image, the total field is $\gamma = \gamma_A + \gamma_R$ where

$$\gamma_A = \frac{60}{r} \mathbf{I} \varepsilon^j \left[\omega t + \frac{\pi}{2} - \frac{2\pi}{\lambda} (r - h \sin \varphi) \right] \quad \dots \quad (24a)$$

$$\gamma_R = - \frac{60}{r} \mathbf{I} \varepsilon^j \left[\omega t + \frac{\pi}{2} - \frac{2\pi}{\lambda} (r + h \sin \varphi) \right] \quad \dots \quad (24b)$$

$$\gamma = \frac{60}{r} \mathbf{I} \varepsilon^j \left(\omega t + \frac{\pi}{2} - \frac{2\pi}{\lambda} r \right) \left(\varepsilon^{j \frac{2\pi}{\lambda} h \sin \varphi} - \varepsilon^{-j \frac{2\pi}{\lambda} h \sin \varphi} \right).$$

$$\text{Now } \left(\varepsilon^{j \frac{2\pi}{\lambda} h \sin \varphi} - \varepsilon^{-j \frac{2\pi}{\lambda} h \sin \varphi} \right) = 2j \sin \left(\frac{2\pi}{\lambda} h \sin \varphi \right)$$

$$\therefore \gamma = \frac{60}{r} \mathbf{I} \varepsilon^j \left(\omega t + \pi - \frac{2\pi}{\lambda} r \right) \times 2 \sin \left(\frac{2\pi}{\lambda} h \sin \varphi \right). \quad \dots \quad (24c)$$

The R.M.S. field being

$$r = \frac{60}{r} I \times 2 \sin \left(\frac{2\pi}{\lambda} h \sin \varphi \right). \quad \dots \quad (25)$$

The factor $2 \sin \left(\frac{2\pi}{\lambda} h \sin \varphi \right)$ may be called the Vertical Distribution Factor and denoted by $D(\varphi)$; it is obviously analogous to the Grating Factor previously used in the case of aerials in free space, the change from "cosine" to "sine" being due merely to the choice of a different datum. The Vertical Distribution Factor is in fact given by the series of polar diagrams in fig. 8,

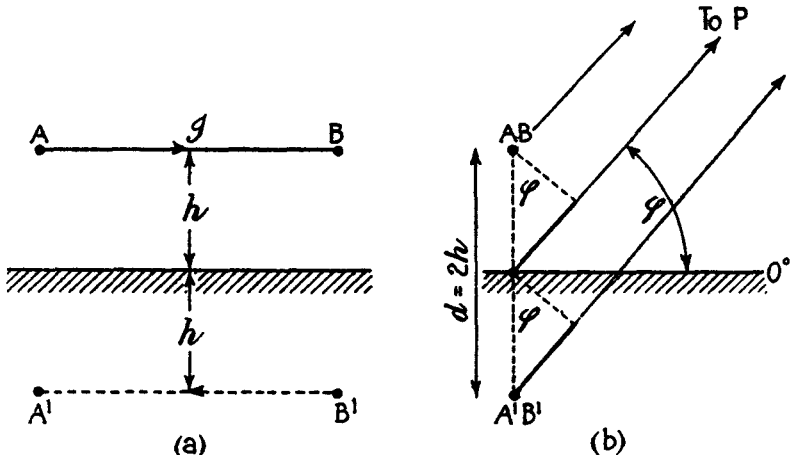


FIG. 20, CHAP. XV.—Notation, horizontal dipole.

line E, except that they must be turned through 90° . In the first few diagrams, one half of the limiting circle has been shaded to represent the ground, so serving as a reminder to perform the necessary rotation. The relation $d = 2h$ must not be forgotten, e.g. the diagram E 18, $d = 4\lambda$, gives $D(\varphi)$ for a height of 2λ .

Effect of height of dipole

38. Several additional vertical polar diagrams are given in fig. 21 (Sheets 1 and 2), and from these it is apparent that no matter what the height may be, there is no radiation along the surface of the earth, even if the latter is perfectly conductive. It follows that with a horizontal aerial, no communication can be performed by means of a true ground ray. If h is less than 0.3λ the greater part of the energy is radiated vertically; if h is increased to about 0.4λ , the diagram shows signs of breaking into two lobes. This kind of diagram is very suitable for a marker beacon in a blind approach system, but for very little else: As h is increased to 0.5λ the two lobes become fully developed, the vertical radiation falling to zero and the maximum field being developed at an angle of 30° to the horizontal. A further increase of height leads to the development of additional lobes, and whenever h is an integral multiple of $\frac{\lambda}{2}$, e.g. λ , $\frac{3}{2}\lambda$, 2λ , etc., the number of lobes is equal to $\frac{4h}{\lambda}$, thus a height of $\frac{5}{4}\lambda$ gives five lobes, and so on. It follows that one lobe is vertical whenever h is an odd multiple of $\frac{\lambda}{4}$ and that no vertical radiation occurs when h is an even multiple of $\frac{\lambda}{4}$. The angles at which successive maxima occur can be determined only

CHAPTER XV.—PARA. 39

approximately from the polar diagrams, but are given to a higher degree of accuracy in fig. 22. The practical use of figs. 21 and 22 is to determine the most effective type of aerial for any particular service. As already stated, heights less than about 0.3λ are useless except for marker

beacons and the like. If $\frac{h}{\lambda}$ is increased to say 0.45 or 0.5 , the aerial may be suitable for short-

distance transmission, e.g. up to about 500 miles, for a projection angle of 45° will give a signal at about that distance by reflection from the F region, assuming the height of the latter to be about 200 miles. For distances of 1,000 miles or more, a projection angle of about 12° to 15° is required; from fig. 21 it is seen that to get maximum radiation at this angle a height of from

λ to $\frac{5}{4}\lambda$ is necessary. Thus if $\lambda = 80$ metres, the radiator must be raised to a height of some 250 to 300 feet.

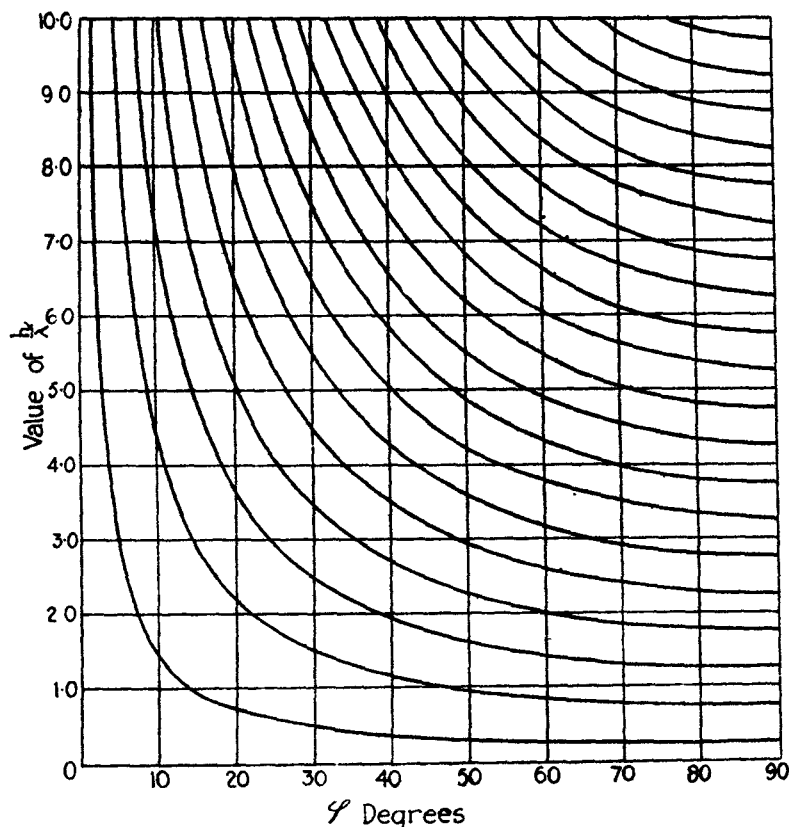
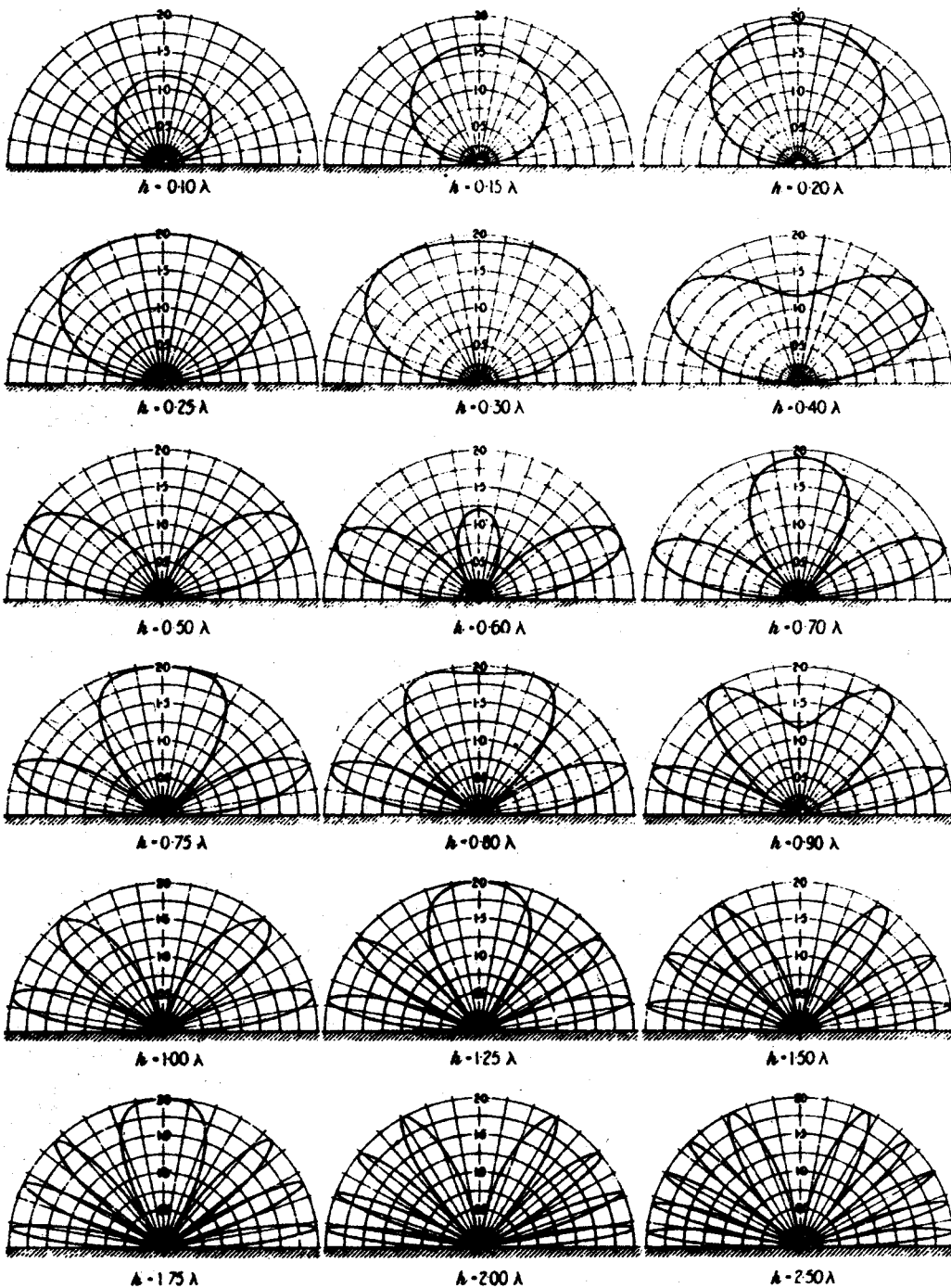


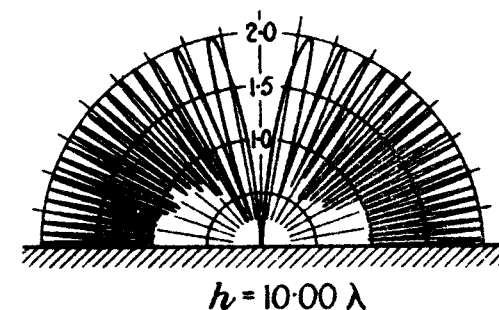
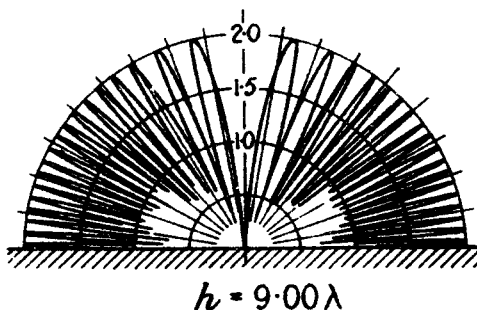
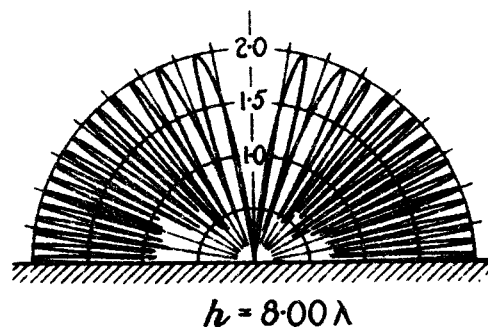
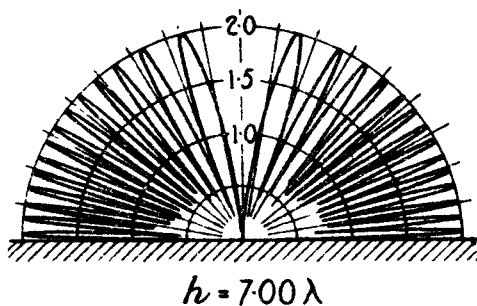
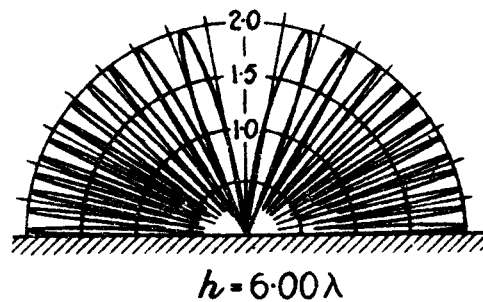
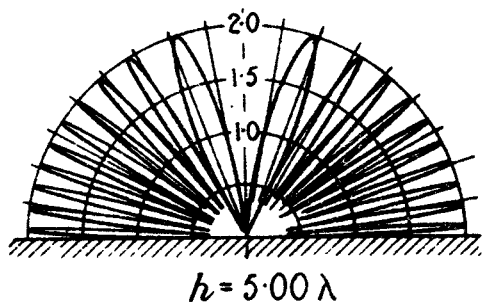
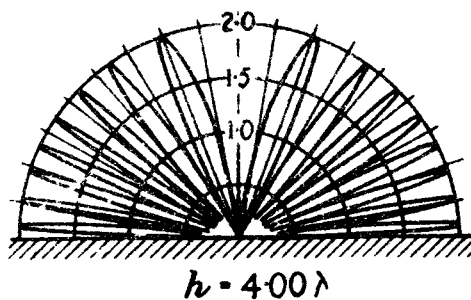
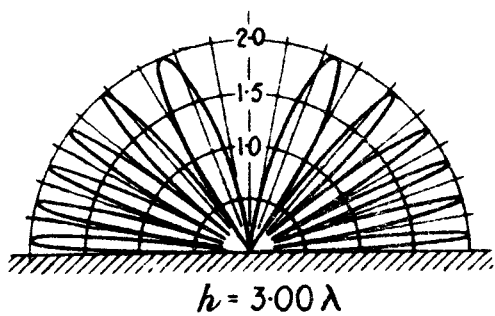
FIG. 22, CHAP. XV.—Horizontal dipole: angles at which field maxima occur.

39. The projection angle of 12° to 15° is suggested for transmission *via* the ionosphere, but where direct-ray communication is necessary, as in ground to air service, a much greater height would be desirable. An aircraft at 10,000 feet and a range of 50 miles is only about 2° above the earth, and if the maximum of the first lobe is to be brought down to 2° , the height of the aerial must be about 7λ . This is of course quite impracticable with a wavelength of 80 metres, and is not very easily or cheaply achieved with so short a wavelength as 3 metres. Nevertheless, it must be regarded as axiomatic that for efficient ground to air communication on high and very



VERTICAL POLAR DIAGRAMS OF HORIZONTAL DIPOLES (SHEET I)

FIG. 21
CHAP. XV.



VERTICAL POLAR DIAGRAMS OF HORIZONTAL DIPOLES
(SHEET 2)

high frequencies, the highest possible masts must be used. Even at comparatively short ranges, the energy received at ground level will generally arrive by reflection from the ionosphere, and the signal will generally be subject to severe fading. When dealing with low angle radiation, i.e. up to a few degrees, the height h , which gives the first maximum at an angle φ_1 radians, is found thus,

$$h \varphi_1 = \frac{\lambda}{4}$$

$$h = \frac{\lambda}{4\varphi_1}$$

If φ_1 is in degrees,

$$h = \frac{14 \cdot 3\lambda}{\varphi_1}$$

Vertical aeriars

40. The vertical polar diagrams of vertical aeriars situated on or near the surface of a perfectly conducting earth are obtained by summing the effects of all the elementary hertzian doublets which comprise the aerial, together with those constituting its image. The elevation of the centre point of the aerial above the ground level is of considerable importance; examples of this, in the case of a $\frac{\lambda}{2}$ dipole, are given qualitatively in Chapter VII. A single example will be given

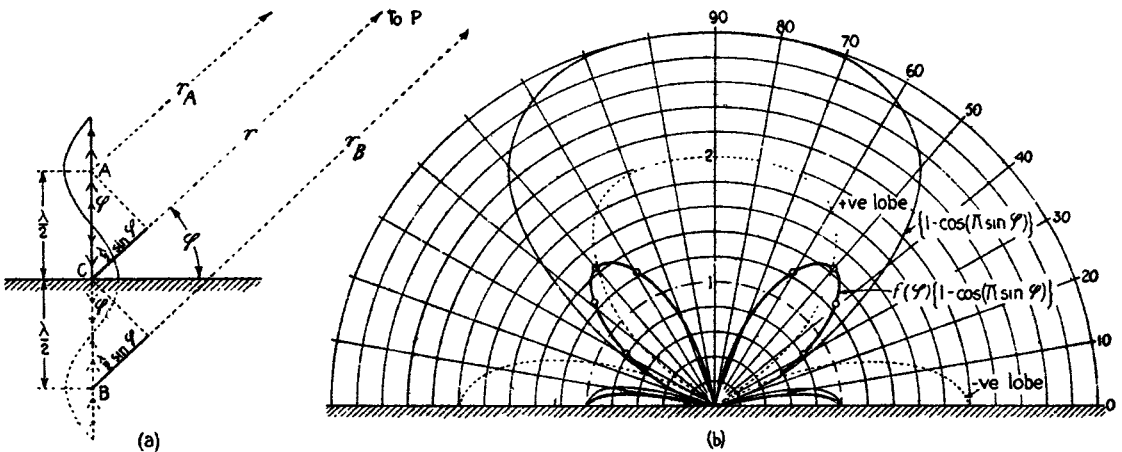


FIG. 23, CHAP. XV.—Calculation of vertical polar diagram of $\frac{3}{4}\lambda$ aerial on perfect earth.

to illustrate the method of calculation. Referring to fig. 23a which shows a $\frac{3}{4}\lambda$ aerial with its lower end at ground level, together with the assumed current distribution in the aerial and its image, it is seen that the length of wire conveniently divides into three $\frac{\lambda}{2}$ sections, A, B, C, each of which may be considered to give rise to a field at the point P. The notation is given in the diagram. The section C gives rise to a field

$$r_0 = \frac{60}{r} f(\varphi) I e^{j\left(\omega t + \frac{\pi}{2} - \frac{2\pi}{\lambda} r\right)} \quad \dots \dots \dots (26)$$

CHAPTER XV.—PARA. 41

Note that $f(\varphi)$ is the appropriate Current Distribution Factor and geometrically is identical with fig. 6, although the angle (φ) is in the present example measured from the ground. The distance of the point P from the current loop of section A is $\left(r - \frac{\lambda}{2} \sin \varphi\right)$, and from the current loop of section B $\left(r + \frac{\lambda}{2} \sin \varphi\right)$. The currents in these sections are in anti-phase to that in section C. The fields due to these sections are therefore

$$\begin{aligned} \gamma_A &= -\frac{60}{r} f(\varphi) I e^{j\left(\omega t + \frac{\pi}{2} - \frac{2\pi}{\lambda} r_A\right)} \\ &= -\frac{60}{r} f(\varphi) I e^{j\left[\omega t + \frac{\pi}{2} - \frac{2\pi}{\lambda} \left(r - \frac{\lambda}{2} \sin \varphi\right)\right]} \dots \dots \dots (27a) \end{aligned}$$

$$\gamma_B = -\frac{60}{r} f(\varphi) I e^{j\left[\omega t + \frac{\pi}{2} - \frac{2\pi}{\lambda} \left(r + \frac{\lambda}{2} \sin \varphi\right)\right]} \dots \dots \dots (27b)$$

and the total field is

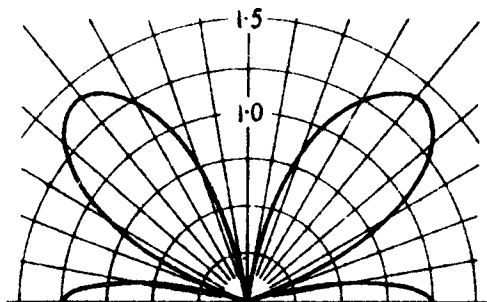
$$\gamma = \gamma_A + \gamma_B + \gamma_C = \frac{60}{r} f(\varphi) I e^{j\left(\omega t + \frac{\pi}{2} - \frac{2\pi}{\lambda} r\right)} \left[1 - \left(e^{+j\pi \sin \varphi} + e^{-j\pi \sin \varphi}\right)\right] \dots (27c)$$

the R.M.S. field being

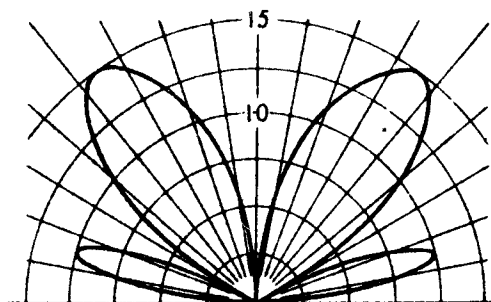
$$\begin{aligned} \Gamma &= \frac{60}{r} f(\varphi) I \left[1 - \left(e^{+j\pi \sin \varphi} + e^{-j\pi \sin \varphi}\right)\right] \\ &= \frac{60}{r} f(\varphi) I \left[1 - 2 \cos (\pi \sin \varphi)\right] \dots \dots \dots (27d) \end{aligned}$$

Vertical polar diagrams

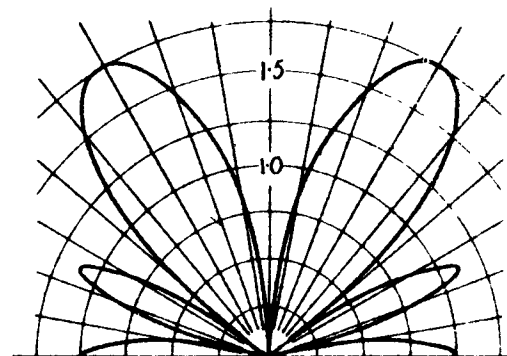
41. The above expression is easily plotted with the aid of fig. 8. Ignoring the term $\frac{60}{r} I$, the shape of the polar diagram can be found thus. Setting aside the factor $f(\varphi)$ for a time, we have to plot $[1 - 2 \cos (\pi \sin \varphi)]$. Now $[2 \cos (\pi \sin \varphi)]$ is given by fig. 8, A 9, turned through 90° , and is shown in dotted line in fig. 23b. Before proceeding further, it is necessary to note that in any diagram of fig. 8 which possesses more than two lobes, the latter are alternatively of positive and negative sign, the lobe extending in the direction 0° being positive. In the dotted line diagram of fig. 23b the signs have been reversed, so that what is shown is $-[2 \cos (\pi \sin \varphi)]$. To this, a circle of radius + 1 unit (shown in chain line) must be added, giving the result shown in thin solid line. The latter diagram represents $[1 - 2 \cos (\pi \sin \varphi)]$, and to obtain the polar diagram, it must be multiplied by the appropriate Current Distribution Factor $f(\varphi)$ (fig. 6) giving the final result shown in heavy line. Proceeding in the above manner the vertical diagrams for vertical aerials of various heights up to 2λ have been calculated, and are shown in fig. 24. It will be observed that all lobes have a common tangent, and further that if the number of quarter wavelengths in the actual aerial is odd, there is always some radiation along the ground, bearing in mind that the latter is assumed to be a perfect conductor. The effect of the finite conductivity will be dealt with later.



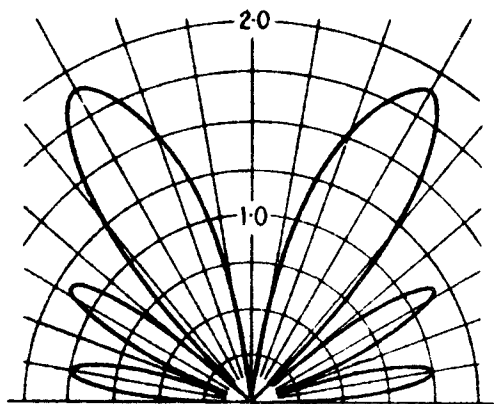
$$h = \frac{3}{4} \lambda$$



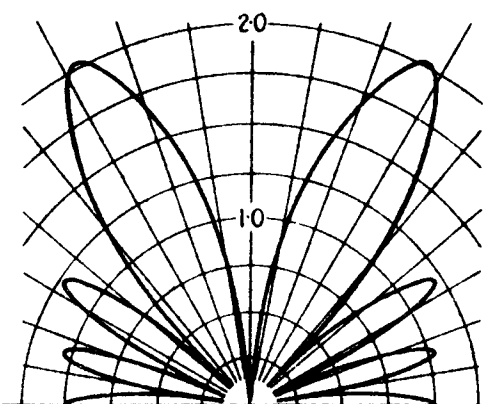
$$h = \lambda$$



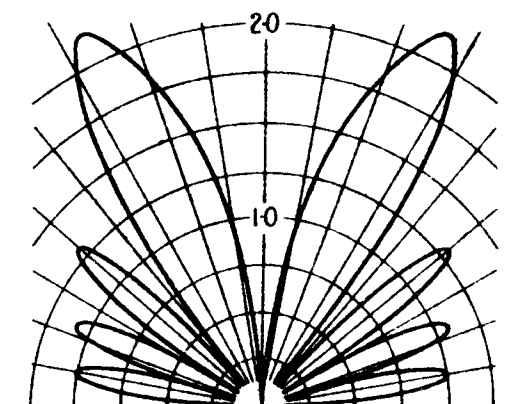
$$h = \frac{5}{4} \lambda$$



$$h = \frac{3}{2} \lambda$$



$$h = \frac{7}{4} \lambda$$



$$h = 2 \lambda$$

Mutual impedance between adjacent radiators

42. Although the elementary theory indicates that certain results will be obtained when given conductors are of particular lengths, or spaced at a particular distance from earth or another conductor, it is found in practice that optimum results are often obtained with slightly different dimensions. In many cases, this is due to the assumption that the mutual impedance between conductors, or between conductor and earth, is zero, whereas it may in fact be of the same order as the impedance of the conductor itself. The magnitude of the mutual impedance between two radiators in free space is defined as the ratio of the induced voltage at the current loop of a second radiator, to the loop current of the first radiator; when dealing with earthed aerials it is convenient to refer the mutual impedance to the base current. The sign of the mutual impedance, when defined in this way, is negative. Thus, if Z_{AB} and Z_{BA} denote the mutual impedances of a radiator A with respect to a radiator B, and *vice versa*, V_{BA} denotes the induced voltage in B due to the current in A. Let I_A be the loop current in A, then

$$Z_{AB} = Z_{BA} = -\frac{V_{BA}}{I_A}.$$

In the following discussion the notation will be as under :—

I_A = vector current in aerial A ; R.M.S. value I_A

I_B = vector current in aerial B ; R.M.S. value I_B

V_A = vector voltage at terminals of aerial A ; R.M.S. value V_A

V_B = vector voltage at terminals of aerial B ; R.M.S. value V_B

$Z_A = \sqrt{R_A^2 + X_A^2}$ = magnitude of self-impedance of aerial A

$$\theta_A = \tan^{-1} \frac{X_A}{R_A}$$

$$z_A = R_A + jX_A = Z_A / \theta_A$$

$Z_B = \sqrt{R_B^2 + X_B^2}$ = magnitude of self-impedance of aerial B

$$\theta_B = \tan^{-1} \frac{X_B}{R_B}$$

$$z_B = R_B + jX_B = Z_B / \theta_B$$

β = phase difference between I_B and I_A

$$M = \frac{I_B}{I_A}, \text{ i.e. } I_B = M I_A / \beta$$

$Z_M = \sqrt{R_M^2 + X_M^2}$ = magnitude of mutual impedance between radiators A and B

$$\theta_M = \tan^{-1} \frac{X_M}{R_M}$$

$$z_M = R_M + jX_M = Z_M / \theta_M$$

R_A' = sum of self and mutual resistance of aerial A

X_A' = sum of self and mutual reactance of aerial A

R_B' = sum of self and mutual resistance of aerial B

X_B' = sum of self and mutual reactance of aerial B

$$z_A' = R_A' + jX_A'$$

$$z_B' = R_B' + jX_B'.$$

CHAPTER XV.—PARAS. 43-44

43. The evaluation of the mutual impedance between two radiators is very tedious and there are very few data available. For identical radiators, the curves of figs. 25 and 26 may be used. These diagrams give the magnitude Z_M and phase angle θ_M for various values of $\frac{d}{\lambda}$. The first-named diagram is applicable to vertical aerials located directly above a perfectly conductive earth, the mutual impedance being referred to the current at the base of the aerial in all cases except for the $\frac{\lambda}{2}$ aerial which is referred to the loop current as usual. Fig. 26a gives Z_M and θ_M for parallel $\frac{\lambda}{2}$ dipoles in free space. It may be noted that the above curves give the self-impedance of the aerial also, i.e. $Z_A / \theta_A = Z_M / \theta_M$ when $\frac{d}{\lambda} = 0$. Fig. 26b may be used for co-linear dipoles. In this case the ordinate, Z_M or θ_M , is plotted against the separation of the adjacent ends. The impedance Z_M / θ_M is easily resolved into its resistive and reactive components by methods previously explained, as in the following.

Example

Find the mutual impedance and reactance between two parallel $\frac{\lambda}{2}$ dipoles A, B, $\frac{\lambda}{2}$ apart in free space.

From fig. 26a, for $\frac{d}{\lambda} = 0.5$

$$Z_M / \theta_M = 33 / -117.$$

$$\therefore R_M = 33 \cos 117 = -33 \cos 63 = -15 \text{ ohms}$$

$$X_M = -33 \sin 117 = 33 \sin 63 = 29.4 \text{ ohms.}$$

From the above example it follows that if the members of such an array are energized by equal, syn-phased currents, the impedance of each is $z + z_M = 73.3 + j42.5 = (15 + j29.4)$ or $58.3 + j13.1$ ohms. The reactance of each member will be annulled by suitable tuning arrangements, while the whole array will have a radiation resistance of 116.6 ohms.

44. Now suppose the two parallel aerials to be energized by voltages V_A, V_B , in such a manner that $I_B = mI_A = M / \beta I_A$. By Kirchoff's laws,

$$V_A = I_A z_A + I_B z_M \quad \dots \dots \dots (28a)$$

$$V_B = I_A z_M + I_B z_B \quad \dots \dots \dots (28b)$$

where z_B may or may not be equal to z_A . Substituting for I_B in equation (28a)

$$\begin{aligned} V_A &= I_A \{R_A + jX_A + m z_M\} \\ &= I_A \{R_A + jX_A + MZ_M / \theta_M + \beta\} \end{aligned}$$

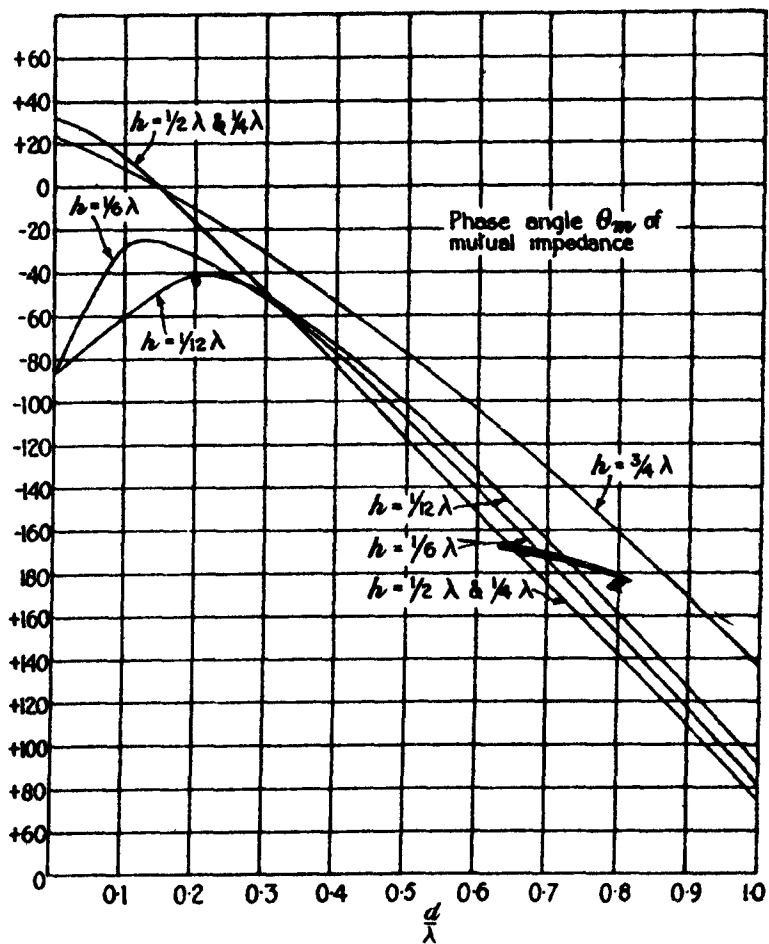
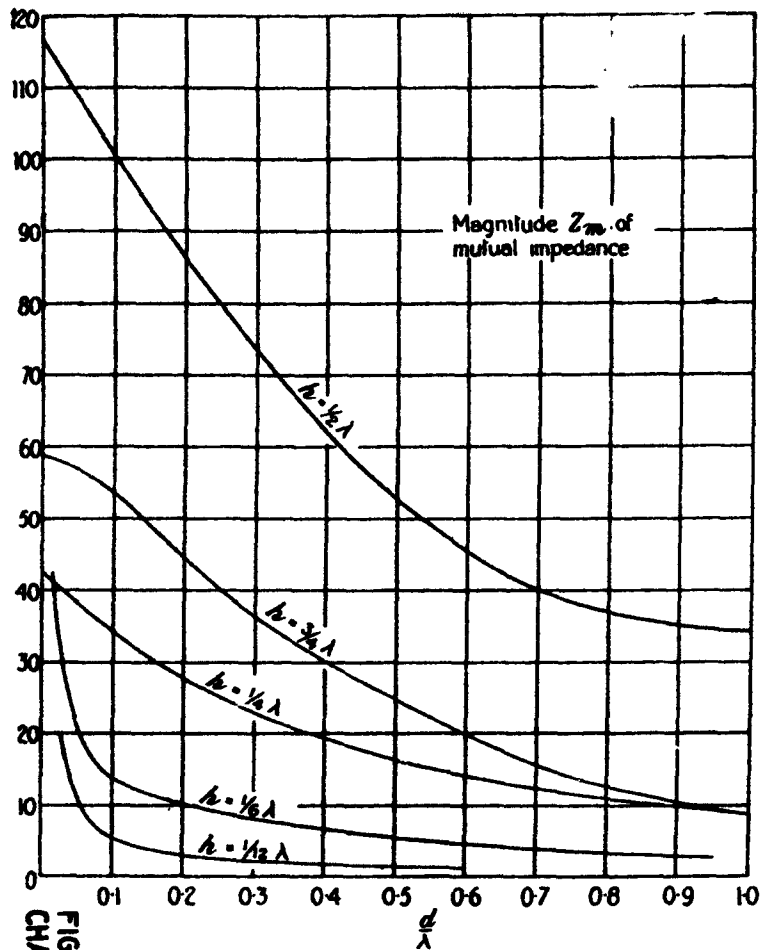
or

$$\begin{aligned} z_A' &= R_A + jX_A + MZ_M \cos(\theta_M + \beta) \\ &\quad + jMZ_M \sin(\theta_M + \beta), \end{aligned}$$

i.e.

$$R_A' = R_A + MZ_M \cos(\theta_M + \beta) \quad \dots \dots \dots (29a)$$

$$X_A' = X_A + MZ_M \sin(\theta_M + \beta). \quad \dots \dots \dots (29b)$$



MUTUAL IMPEDANCE Z_m $\angle \theta_m$ BETWEEN IDENTICAL VERTICAL AERIALS

Similarly

$$R_B' = R_B + \frac{Z_M}{M} \cos (\theta_M - \beta) \quad \dots \dots \dots (29c)$$

$$X_B' = X_B + \frac{Z_M}{M} \sin (\theta_M - \beta). \quad \dots \dots \dots (29d)$$

It is now seen that the radiation resistances of the two aerials are only equal when $M = 1$ and β is either 0° or 180° . It is of interest to note that in any other circumstances the reactances are also unequal, so that the tuning reactances of the two aerials will differ.

Effect upon power distribution

45. As an example of the effects of the value of m , let us consider two $\frac{\lambda}{4}$ aerials, A and B, $\frac{\lambda}{4}$ apart upon a perfect earth, and energized with currents I_A, I_B , where $I_B = 0.8 I_A / 90^\circ$. The self impedance of each aerial will be $36.6 + 21.5$ ohms. From fig. 25, $z_M = 25 / -36$ ohms. Then

$$\begin{aligned} R_A' &= R_A + MZ_M \cos (90 - 36) \\ &= 36.6 + 0.8 \times 25 \cos 54 \\ &= 36.6 + 0.8 \times 25 \times 0.588 \\ &= 36.6 + 11.76 \\ &= 48.36 \text{ ohms} \end{aligned}$$

$$\begin{aligned} R_B' &= R_B + \frac{Z_M}{M} \cos (-90 - 36) \\ &= 36.6 + \frac{25}{0.8} \cos (-126) \\ &= 36.6 - \frac{25}{0.8} \times 0.588 \\ &= 36.6 - 18.4 \\ &= 18.2 \text{ ohms.} \end{aligned}$$

Let us now find how these aerials would share a power of 500 watts.

$$I_A^2 R_A' + I_B^2 R_B' = 500$$

$$I_A^2 (48.36 + 0.8^2 \times 18.2) = 500$$

$$I_A^2 \times 60 = 500$$

$$I_A = \sqrt{\frac{500}{60}}$$

$$= 2.9 \text{ amperes nearly}$$

$$I_B = 0.8 I_A = 2.32 \text{ amperes}$$

$$\begin{aligned} P_A &= I_A^2 R_A' = 2.9^2 \times 48.36 \\ &= 405 \text{ watts} \end{aligned}$$

$$\begin{aligned} P_B &= I_B^2 R_B' = 2.32^2 \times 17.2 \\ &= 93 \text{ watts} \end{aligned}$$

$$P_T = 498 \text{ watts,}$$

CHAPTER XV.—PARA. 46

the slight discrepancy being due to arithmetical approximations. In order to energize the aerials in the manner specified, the applied voltages V_A , V_B must differ both in phase and magnitude.

The vector ratio $\frac{V_A}{V_B}$ is easily found

$$\begin{aligned} V_A &= I_A z_A + m I_A z_M \\ V_B &= I_A z_M + m I_A z_B \\ \therefore \frac{V_A}{V_B} &= \frac{z_A + m z_M}{z_M + m z_B} \end{aligned} \quad \dots \dots \dots (30)$$

This equation determines the nature of the network which must separate the feeding points of A and B, if they are to be supplied from the same feeder line.

Effect upon polar diagram

46. It is also of importance to appreciate the extent to which the polar diagrams of an array are affected by the mutual impedance. Taking the two aerials A and B of the previous paragraph, but separated by an unspecified distance d , the total power radiated will be $I_A^2 R_A' + I_B^2 R_B'$.

Putting $I_B = M/\beta I_A$ as before, the total power radiated is

$$P_T = I_A^2 (R_A + M^2 R_A + 2M R_M \cos \beta).$$

Thus, for a given power P_T the current in A is

$$I_A = \sqrt{\frac{P_T}{R_A \left(1 + M^2 + 2M \frac{R_M}{R_A} \cos \beta\right)}}$$

In paragraph 30 the polar diagram of two parallel radiators is derived on the assumption that $R_M = 0$. It is then shown that at any point P having the co-ordinates r , θ , φ , the R.M.S. field is

$$r = \frac{60}{r} F.f(\varphi) \sqrt{\frac{P_T}{R_A (1 + M^2)}} \sqrt{1 + M^2 + 2M \cos \left(\beta + \frac{2\pi}{\lambda} d \cos \varphi \cos \theta \right)} \quad \dots \dots \dots (31)$$

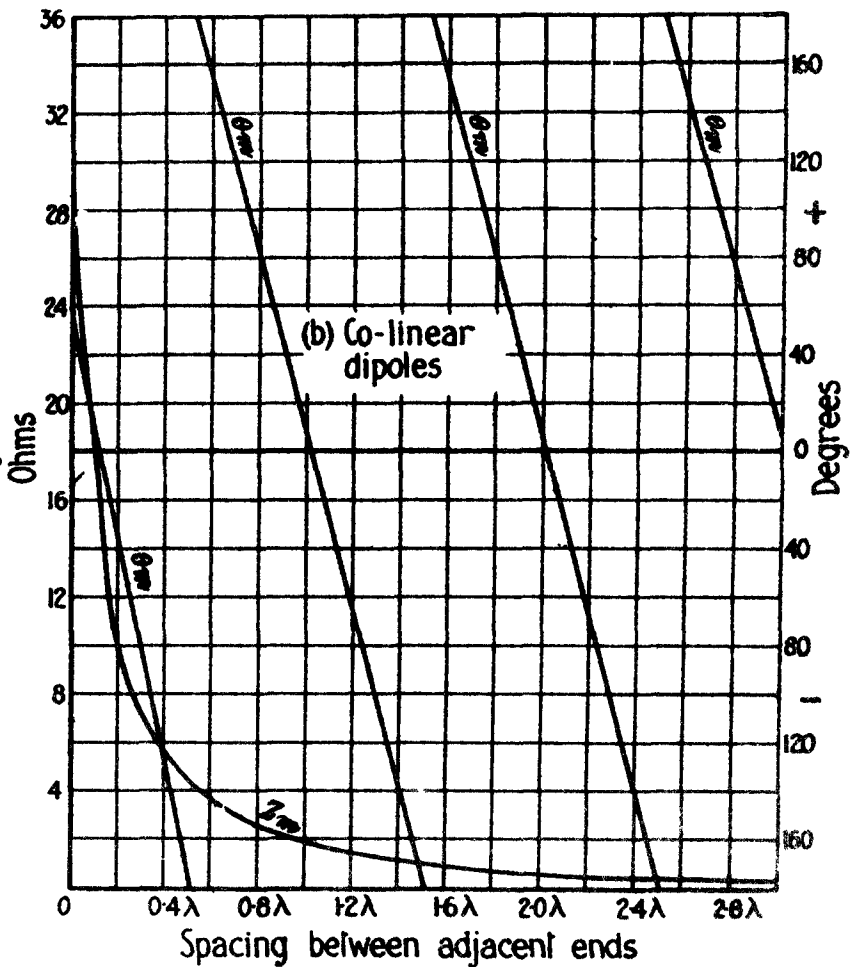
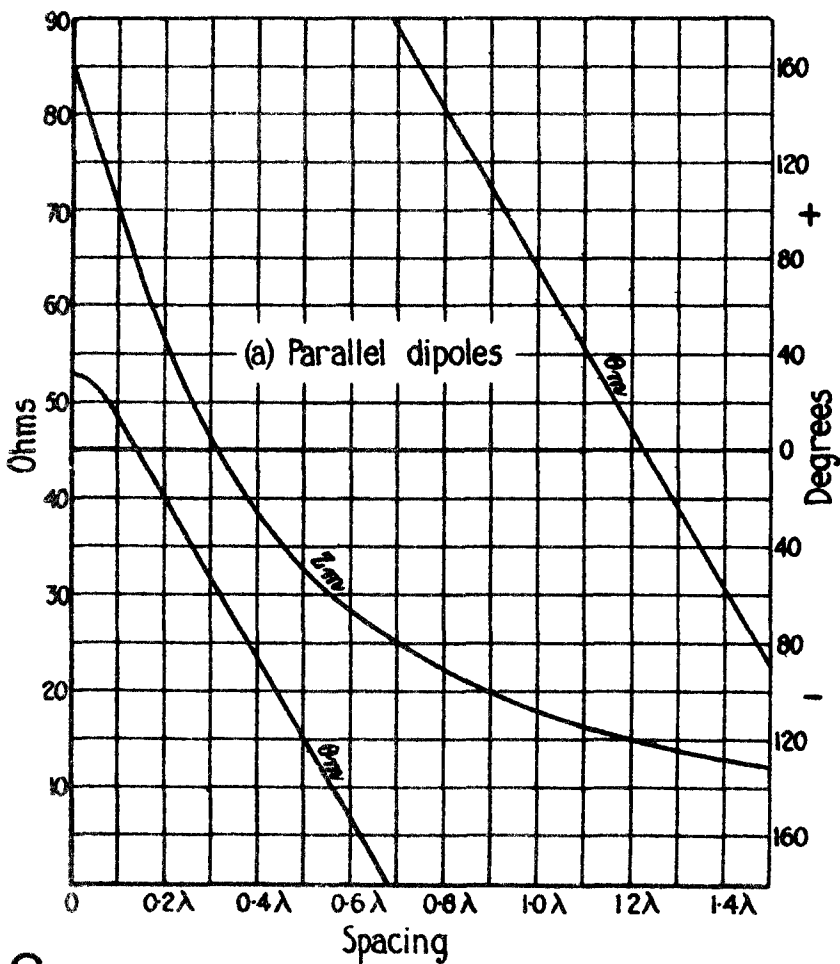
where $\sqrt{\frac{P_T}{R_A (1 + M^2)}}$ is the current in aerial A.

To allow for the mutual impedance, then, we have only to substitute the value of I_A as modified by the mutual resistance R_M , giving

$$r = \frac{60}{r} F.f(\varphi) \sqrt{\frac{P_T \left[1 + M^2 + 2M \cos \left(\beta + \frac{2\pi d}{\lambda} \cos \varphi \cos \theta \right) \right]}{R_A \left(1 + M^2 + 2M \frac{R_M}{R_A} \cos \beta \right)}} \quad \dots \dots \dots (32)$$

The horizontal polar diagram is obtained by putting $\varphi = 0$, $\cos \varphi = 1$. The field at ground level is then

$$r_0 = \frac{60}{r} F \sqrt{\frac{P_T \left[1 + M^2 + 2M \cos \left(\beta + \frac{2\pi d}{\lambda} \cos \theta \right) \right]}{R_A \left(1 + M^2 + 2M \frac{R_M}{R_A} \cos \beta \right)}} \quad \dots \dots (33)$$



MUTUAL IMPEDANCE OF DIPOLES IN FREE SPACE

If the same power were supplied to the aerial A alone, B being entirely removed, the polar diagram would be a circle, the field strength being given by

$$\Gamma_{0(A)} = \frac{60}{r} F \sqrt{\frac{P_T}{R_A}},$$

from paragraph 14. The ratio of Γ_0 to $\Gamma_{0(A)}$ is

$$\frac{\Gamma_0}{\Gamma_{0(A)}} = \sqrt{\frac{1 + M^2 + 2M \cos \left(\beta + \frac{2\pi}{\lambda} d \cos \theta \right)}{1 + M^2 + 2M \frac{R_M}{R_A} \cos \beta}}. \quad \dots \dots (34)$$

47. From this ratio it is easy to plot polar diagrams corresponding to those of fig. 8, but it is obviously impossible to portray all the possibilities. For the particular case when $M = 1$, $R_M \neq 0$, we have a further simplification

$$\begin{aligned} \frac{\Gamma_0}{\Gamma_{0(A)}} &= \sqrt{\frac{1 + \cos \left(\beta + \frac{2\pi}{\lambda} d \cos \theta \right)}{1 + \frac{R_M}{R_A} \cos \beta}} \\ &= \frac{\sqrt{2} \cos \left(\frac{\beta}{2} + \frac{\pi d}{\lambda} \cos \theta \right)}{\sqrt{1 + \frac{R_M}{R_A} \cos \beta}} \end{aligned}$$

The numerator of this expression obviously gives the ratio $\frac{\Gamma_0}{\Gamma_{0(A)}}$ if $R_M = 0$, as in paragraph 31.

Since neither $\frac{R_M}{R_A}$ nor $\cos \beta$ can exceed unity, the product $\frac{R_M}{R_A} \cos \beta$ cannot exceed unity and may be very much less. Thus, with equal currents in both aerials, the shape of the polar diagram is very little affected by the mutual impedance, so that fig. 8 may be used for practical purposes even though the mutual impedance was not taken into account in calculating the diagrams.

Effect of Z_M upon current in radiating members

48. (i) In an array consisting of more than two radiating members, the mutual impedance between the radiators may exercise a considerable influence upon the radiation characteristics. As an illustration the array shown in fig. 27 will be considered briefly. Here A, B, and C are parallel $\frac{\lambda}{2}$ aerials on a perfectly conductive earth, and are spaced $\frac{\lambda}{2}$ apart. Each aerial has a self-impedance $z = R + jX$ ohms; tuning reactances X_1, X_2, X_3 , may or may not be included. Let $z_A = z + jX_1, z_B = z + jX_2, z_C = z + jX_3$. Let us first assume that the tuning reactances are absent, and that the aerials are fed from a common source of voltage V . We then have

$$V = zI_A + z_I I_B + z_C I_C \quad \dots \dots \dots (35a)$$

$$V = z_I I_A + zI_B + z_C I_C \quad \dots \dots \dots (35b)$$

$$V = z_C I_A + z_I I_B + zI_C \quad \dots \dots \dots (35c)$$

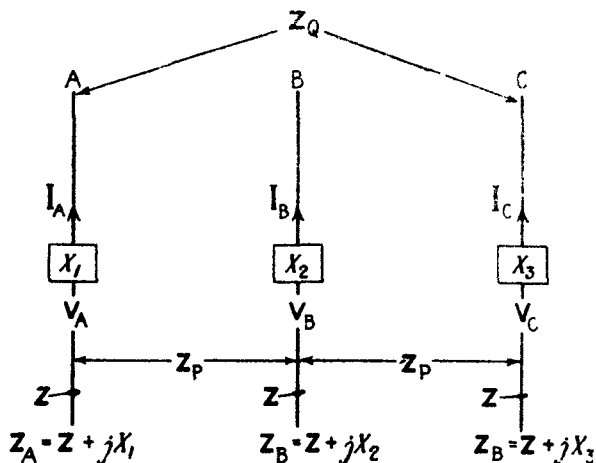


FIG. 27, CHAP. XV.—Notation—three parallel aerials.

From the symmetry of the arrangement it is obvious that $I_A = I_C$, although I_B is not necessarily equal to I_A . Let $I_B = mI_A$. Then

$$V = (z + z_Q + mz_P) I_A$$

$$V = (2z_P + mz) I_A$$

$$z + z_Q + mz_P = 2z_P + mz$$

$$\text{and } m = \frac{z + z_Q - 2z_P}{z - z_P} \quad \dots \dots \dots (36)$$

This ratio is easily evaluated. From fig. 25 the various impedances are found to be (to the nearest integers)

$$z = 100 + j58$$

$$z_P = - (24 + j47)$$

$$z_Q = 10 + j32$$

$$\therefore m = \frac{100 + j58 + 10 + j32 + 48 + j94}{100 + j58 + 24 + j47}$$

$$= \frac{158 + j184}{124 + j105}$$

$$= 1.43 + j0.277$$

$$= 1.46 / 11^\circ \text{ approximately.}$$

Thus the current in the centre aerial is nearly 50 per cent. greater than in the outer ones and is slightly out of phase. For practical purposes the horizontal polar diagram may be obtained by adding a circle of radius 1.46 units to diagram A 9 of fig. 8, taking the vertical lobes to be of positive and the horizontal lobes to be of negative sign, and ignoring the effect of the slight phase difference.

(ii) Next we shall suppose the aerials to be individually resonant, so that $z_A = z_B = z_C = R = 100$ ohms. From the previous example it is easily seen that in this case

$$\begin{aligned} m &= \frac{R + z_Q - 2z_F}{R - z_F} \\ &= \frac{100 + 10 + j32 + 48 + j94}{100 + 24 + j47} \\ &= \frac{158 + j126}{124 + j47} \\ &= 1.49 / 18^\circ \end{aligned}$$

Thus the current in the centre aerial is still nearly fifty per cent. greater than in the outer ones, and is out of phase by a greater angle than before.

49. Finally, let us find the conditions under which $m = 1 / 0^\circ$, the current in each aerial to be in phase with its supply voltage. To achieve this it will be necessary to feed the centre aerial in such a manner that V_B is not equal to V_A .

$$V_A = z_A I_A + z_F I_B + z_Q I_C \quad \dots \dots \dots (37a)$$

$$V_B = z_F I_A + z_B I_B + z_F I_C \quad \dots \dots \dots (37b)$$

But $I_A = I_B = I_C$;

$$V_A = (z + jX_1 + z_F + z_Q) I_A \quad \dots \dots \dots (38a)$$

$$V_B = (2z_F + z + jX_2) I_B \quad \dots \dots \dots (38b)$$

i.e.

$$z_A' = R + jX + jX_1 + R_F + jX_F + R_Q + jX_Q$$

$$\begin{aligned} R_A' &= R + R_F + R_Q \\ &= 100 + 24 + 10 \\ &= 86 \text{ ohms} \end{aligned}$$

$$\begin{aligned} X_A' &= X + X_1 + X_F + X_Q = 0 \\ X_1 &= -(X + X_F + X_Q) \\ &= -(58 - 47 + 32) \\ &= -43 \text{ ohms} \end{aligned}$$

$$\begin{aligned} z_B' &= 2R_F + 2jX_F + R + jX + jX_2 \\ R_B' &= R + 2R_F \\ &= 100 + 48 \\ &= 148 \text{ ohms.} \end{aligned}$$

$$\begin{aligned} X_B' &= X + X_2 + 2X_F = 0 \\ X_2 &= -(X + 2X_F) \\ &= -(58 - 94) \\ &= 36 \text{ ohms.} \end{aligned}$$

Thus in order to tune the array correctly, it is necessary to insert capacitive reactances in the outer members and an inductive reactance in the centre member. In practice this tuning may be achieved by suitable adjustment of the length of the radiator, or by the addition of a susceptance in parallel with the aerial instead of a series reactance. Such susceptances may take the form of

CHAPTER XV.—PARA. 50

short lengths of non-radiating feeder line. It is obviously necessary to supply the outer aerials at a higher voltage than the centre one, i.e. $V_A = V_C = 1.65 V_B$. Since the aerials are carrying equal currents, we may refer to the radiation resistance of the whole aerial without ambiguity. This is equal to $R_A' + R_C' + R_B' = 208$ ohms, the average resistance per radiator being 69 ohms. It is found that as the number of parallel syn-phased dipoles is increased, the average resistance per aerial falls slightly, approaching about 56 ohms for an infinite number of aerials. On the other hand, if $\frac{\lambda}{2}$ aerials, spaced $\frac{\lambda}{2}$ apart, are fed with equal currents having a progressive phase difference of 180° , in order to obtain an "end-fire" diagram, a repetition of the above calculation gives the total radiation resistance of these aerials as 416 ohms, an average of 139 ohms per radiator.

Effect of Z_M between radiator and reflector

50. Let us now consider the action of a reflector aerial more fully than in paragraphs 33 and 34. In the notation previously used let A and B be two parallel vertical aerials separated by a distance d , A being energized by the application of an oscillatory voltage V , and B being unenergized. Let their self-impedance operators be z_A , z_B and their mutual impedance operator z_M . By Kirchoff's law

$$V = z_A I_A + z_M I_B \quad \dots \dots \dots (39a)$$

$$0 = z_M I_A + z_B I_B \quad \dots \dots \dots (39b)$$

The current in the reflector B is

$$I_B = -\frac{z_M}{z_B} I_A$$

$$= m I_A.$$

Thus

$$m = -\frac{z_M}{z_B} = -\frac{Z_M / \theta_M}{Z_B / \theta_B}$$

$$= \frac{Z_M}{Z_B} / \pi + \theta_M - \theta_B$$

$$V = \left\{ z_A - \frac{(z_M)^2}{z_B} \right\} I_A$$

$$\frac{V}{I_A} = z_A' = z_A - \frac{(z_M)^2}{z_B}$$

$$= R_A + jX_A - \frac{(Z_M)^2}{Z_B} / 2\pi + 2\theta_M - \theta_B$$

$$= R_A + jX_A - \frac{(Z_M)^2}{Z_B} / 2\theta_M - \theta_B$$

so that

$$R_A' = R_A - \frac{(Z_M)^2}{Z_B} \cos (2\theta_M - \theta_B) \quad \dots \dots \dots (40a)$$

$$X_A' = X_A - \frac{(Z_M)^2}{Z_B} \sin (2\theta_M - \theta_B). \quad \dots \dots \dots (40b)$$

51. As would be expected, then, the presence of the reflector modifies both the resistance and the reactance of the energized aerial. The polar diagram is calculated by methods already explained. If F_0 is the field strength in the horizontal plane in the direction θ when the reflector is absent, and F_1 the field in the same direction with the reflector present, for the same input power,

$$\begin{aligned} F_1 &= F_0 \sqrt{\frac{R_A}{R_A'} \left\{ 1 + M^2 + 2M \cos \left(\beta - \frac{2\pi}{\lambda} d \cos \theta \right) \right\}} \\ &= F_0 \sqrt{\frac{1 + M^2 + 2M \cos \left(\beta - \frac{2\pi}{\lambda} d \cos \theta \right)}{1 - M \frac{Z_M}{R_A} \cos (2\theta_M - \theta_B)}} \quad \dots \quad (41) \end{aligned}$$

A study of the above results will show that it is almost if not quite impossible to fulfil the conditions required to give a horizontal polar diagram corresponding exactly with fig. 8 C 3. To obtain the latter diagram it is necessary to have $m = 1 / \frac{\pi}{2}$. Now $M = \frac{Z_M}{Z_B}$ and Z_B cannot be less than R_B . If the aerials are $\frac{\lambda}{2}$ dipoles on a perfectly conductive earth, the minimum value of R_B is 100 ohms. The current in the reflector aerial leads on that in the energized aerial by an angle $\beta = 180^\circ + \theta_M - \theta_B$, and if $Z_B = R_B$, $\theta_B = 0$. Thus the mutual impedance should have a phase angle of -90° . Reference to fig. 25 shows that $\theta_M = -90^\circ$ when the spacing is approximately 0.4λ and Z_M is then only 60 ohms, so that $m = 0.6 / \frac{\pi}{2}$ instead of $1 / \frac{\pi}{2}$.

52. The above position may be summarized by the statement that it is impossible simultaneously to fulfil the conditions that the forward radiation shall be double that of the single energized aerial, and the backward radiation absolutely annulled, by the use of a single reflector aerial. So far as it is possible to generalize, it may be said that in the case of a single aerial with reflector, both tuned to the same frequency, the optimum spacing for maximum forward radiation is approximately $\frac{\lambda}{2\pi}$, and for minimum backward radiation, $\frac{\lambda}{\pi}$. For the optimum ratio of forward to backward radiation, the separation should be about 0.28λ . These results are only of practical importance when a single energized member and a single reflector are used. When reflector aerials are used in conjunction with arrays consisting of several radiating members, the spacing is not critical, and it is found that a spacing of $\frac{\lambda}{4}$ is as effective as any, the reflector being usually slightly mistuned as explained below. In certain designs, particularly on the higher frequencies, a spacing of $\frac{3}{4}\lambda$ is sometimes adopted.

Mistuning of reflector

53. It is possible to obtain a near approach to the desired cardioid diagram by mistuning the reflector aerial, the degree of mistuning being dependent upon the spacing; when this expedient is adopted the $\frac{\lambda}{4}$ spacing is in most circumstances as effective as any other. For any given set of conditions, the horizontal polar diagram is easily calculated from the expressions given above, particularly since, if only the shape of the diagram is required, it is sufficient to plot the portion

CHAPTER XV.—PARA. 54

$\sqrt{1 + M^2 + 2M \cos(\beta - \frac{2\pi}{\lambda} d \cos \theta)}$. The method may be seen from the following example.

If A and B are vertical $\frac{\lambda}{2}$ aerials, $\frac{\lambda}{4}$ apart upon a conductive earth, A only being energized, the radiation resistance of each wire will be 73 ohms and the dead-loss resistance may be only 2 ohms, so that the total resistance of each is 75 ohms. It is not suggested that the dead-loss resistance can be kept within so low a figure in practice, but it will be seen that unless the dead-loss resistance is very low the desired diagram cannot be obtained.

54. For $\frac{d}{\lambda} = 0.25$, fig. 25 gives Z_M / θ_M as $80 / -35$. Suppose the reflector to be mistuned, having a positive, i.e. (inductive) reactance, its impedance being $Z_B / 45^\circ$.

Then
$$Z_B = \frac{R_B}{\cos \theta_B} = \frac{75}{0.707} = 106 \text{ ohms}$$

and
$$M = \frac{Z_M}{Z_B} = \frac{80}{106} = 0.755$$

$$M^2 = 0.57$$

The angle β by which the current I_B leads on I_A is given by

$$\begin{aligned} \beta &= \pi + \theta_M - \theta_B = (180 - 35 - 45) \text{ degrees} \\ &= 100^\circ \end{aligned}$$

Substituting these values of M and β in the expression $1 + M^2 + 2M \cos(\beta - \frac{2\pi}{\lambda} d \cos \theta)$ we obtain $1.57 + 1.51 \cos(100 - 90 \cos \theta)$; when $\theta = 0$, $\cos \theta = 1$, $(100 - 90 \cos \theta) = 10$.

$$1.51 \cos 10 = 1.486$$

$$1.57 + 1.486 = 3.056$$

Hence the field in this direction is

$$F_0 = F_A \sqrt{\frac{R_A}{R_A'}} \sqrt{3.056}.$$

Ignoring the terms $F_A \sqrt{\frac{R_A}{R_A'}}$, which give the scale of the diagram

$$\begin{aligned} F_0 &= \sqrt{3.056} \\ &= 1.75. \end{aligned}$$

When

$$\theta = 180, \cos \theta = -1, (100 - 90 \cos \theta) = 190$$

$$1.51 \cos 190 = -1.51 \cos 10 = -1.486$$

$$1.57 - 1.486 = 0.084$$

$$\begin{aligned} F_{180} &= \sqrt{0.084} \\ &= 0.29 \end{aligned}$$

The field in other directions is found in the same manner and so the shape of the horizontal polar diagram is determined. Actually a good approximation may be found by calculating F_0 and F_{180} as above, and in addition, F_{90} and the minimum field. The field F_{90} is obviously

$\sqrt{1 + M^2 + 2M \cos \beta}$; in the given example this becomes $\sqrt{1.57 + 1.51 \cos 100^\circ} = \sqrt{1.57 - 1.51 \times 0.1736} = 1.15$. The minimum field obviously occurs when

$\cos\left(\beta - \frac{2\pi}{\lambda} d \cos \theta\right) = -1$, i.e. when $\left(\beta - \frac{360d}{\lambda} \cos \theta\right) = 180$. In the present instance we have

$$100 - 90 \cos \theta = 180$$

$$\cos \theta = -\frac{8}{9}$$

$$\theta = 153^\circ$$

The field F_{153} is equal to $\sqrt{1 + M^2 - 2M} = \sqrt{1.57 - 1.51} = 0.245$.

55. With regard to the scale, we have to find F_A , the field which would be set up by the aerial A alone. This is equal to $\frac{60}{r} I_A f(\varphi)$. Geometrically $f(\varphi)$ is identical with fig. 6 and its magnitude in the horizontal plane ($\varphi = 0$) is unity. Next the expression $\sqrt{\frac{R_A}{R_A'}}$ must be evaluated. From equation 40a,

$$\begin{aligned} \frac{R_A'}{R_A} &= 1 - \frac{Z_M^2}{Z_B R_A} \cos(2\theta_M - \theta_B) \\ &= 1 - M \frac{Z_M}{R_A} \cos(2\theta_M - \theta_B). \end{aligned}$$

In the present example this becomes

$$\begin{aligned} \frac{R_A'}{R_A} &= 1 - \frac{0.755 \times 80}{73} \cos(-70 - 45) \\ &= 1.35 \\ \frac{R_A}{R_A'} &= 0.74 \\ \sqrt{\frac{R_A}{R_A'}} &= 0.86. \end{aligned}$$

Finally, allowance must be made for the radiation contributed by the virtual image of the array. This entails the introduction of a Vertical Distribution Factor $D(\varphi)$; since the centre point of the array is $\frac{\lambda}{4}$ above the earth, the appropriate factor is geometrically identical with fig. 8 A 5, but must be turned through 90° so that it has the value 2 along the ground ($\varphi = 0, \theta = 0$). Thus the R.M.S. field in the direction $\theta = 0$, along a perfectly conductive ground will be $2 \times 0.86 \times 1.75 = 3 \times \frac{60}{r} I$ and in the direction $\theta = 180, 2 \times 0.86 \times 0.29 = 0.465 \times \frac{60}{r} I$.

The complete horizontal polar diagram is therefore that shown in fig. 28b. From the manner in which it is obtained it is obvious that if the diagram is rotated through 90° about the axis XX, and then multiplied by the appropriate values of $f(\varphi)$ and $D(\varphi)$ as defined above, the vertical polar diagram (fig. 28c) is obtained. A few points so calculated will give a sufficiently close approximation.

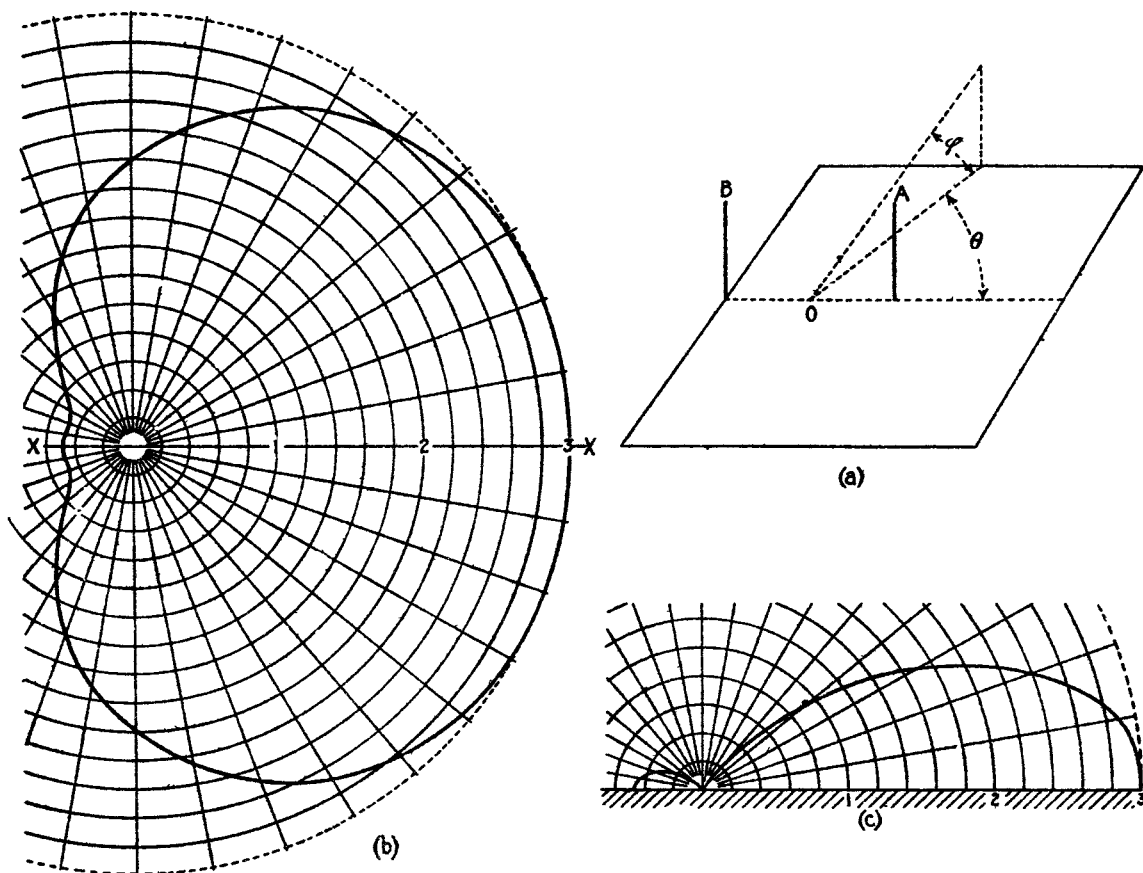


FIG. 28. CHAP. XV.—Example of calculation of polar diagrams.

Effect of dead-loss resistance

56. The importance of low dead-loss resistance in the reflector aerial can easily be appreciated. Again referring to fig. 8, which, it will be remembered, is constructed on the basis of equal currents in the two aerials, it is seen that if this condition is fulfilled, an approach to the desired unidirectional diagram is attainable even if the respective currents are not in quadrature. For example, compare diagrams B 3, B 4, D 2, D 3, with diagram C 3. If the two currents are not equal, however, the attainment is much more difficult, and it is therefore desirable to make M approach unity as closely as possible. Since $M = \frac{Z_{\text{A}}}{Z_{\text{B}}}$ and Z_{A} is constant for any particular spacing, M can only approach unity if Z_{B} is kept small. Even with zero dead-loss resistance and zero resistance, Z_{B} is equal to the radiation resistance, e.g. for a $\frac{\lambda}{2}$ dipole, Z_{B} must be more than 73 ohms. This considerably narrows the range of $\frac{d}{\lambda}$ from which a suitable value of Z_{A} can be chosen.

Influence of finite conductivity and permittivity of ground

57. We may now briefly discuss the errors involved in the assumption that the surface of the earth is a perfect conductor, so far as its properties as a reflector are concerned. Since we are only concerned with the earth in the vicinity of the aerial, we shall consider the surface to be plane as before. Fresnel's equations governing the reflection of a plane electro-magnetic wave at a plane surface are given in the previous chapter, in terms of the angle of incidence as defined for physical purposes. For the present purpose, however, it is more convenient to state them in terms of the ground angle as previously used in this Chapter. The ratio of the field strength on reflection, Γ_r , to the incident field strength Γ_i , is then a complex number. Two different solutions occur according to the plane of polarization of the incident wave. For a wave polarized in the plane of incidence, i.e. vertical polarization, we have

$$\frac{\Gamma_r}{\Gamma_i} = K_v / \theta_v = K_v e^{j\theta_v} = k_v,$$

while if the wave is polarized perpendicularly to the plane of incidence, i.e. horizontal polarization,

$$\frac{\Gamma_r}{\Gamma_i} = K_h / \theta_h = K_h e^{j\theta_h} = k_h.$$

In terms of the ground angle φ , Fresnel's equations become

$$k_v = \frac{\left(\kappa - j \frac{2\sigma}{f}\right) \sin \varphi - \sqrt{\kappa - \cos^2 \varphi - j \frac{2\sigma}{f}}}{\left(\kappa - j \frac{2\sigma}{f}\right) \sin \varphi + \sqrt{\kappa - \cos^2 \varphi - j \frac{2\sigma}{f}}} \quad \dots \quad (42)$$

$$k_h = \frac{\sqrt{\kappa - \cos^2 \varphi - j \frac{2\sigma}{f}} - \sin \varphi}{\sqrt{\kappa - \cos^2 \varphi - j \frac{2\sigma}{f}} + \sin \varphi} \quad \dots \quad (43)$$

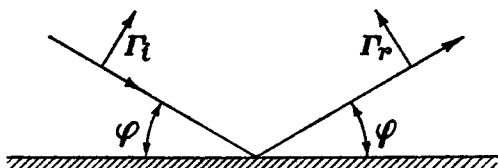
where κ is the permittivity of the ground.

σ is the conductivity of the ground in E.S.U.

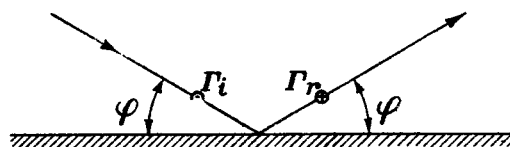
φ is the ground angle, i.e. the complement of the angle of incidence.

f is the frequency in cycles per second.

58. The expressions k_v , k_h are referred to as the complex coefficients of reflection for the respective cases. Their moduli, K_v and K_h , are always less than unity. The angle θ_v or θ_h must be added to the phase of the incident wave to obtain the phase of the reflected wave. This angle is always negative, and lies between 0 and -180° . In using these equations it is most important to observe the conventions which have been adopted in obtaining them. These are shown diagrammatically in fig. 29. Taking the vertical polarization case first, Γ_i and Γ_r are both considered to be positive in the upward direction, along the plane of the paper. In the case of



(a) Γ in plane of incidence



(b) Γ perpendicular to plane of incidence

FIG. 29, CHAP. XV.—Conventional positive directions of electric field vector.

CHAPTER XV.—PARAS. 59-60

horizontally polarized waves E_h is considered to be positive when it is in a direction upward from the surface of the paper, and the positive direction of E_v is considered to be downward, i.e. below the surface of the paper. The importance of these conventions becomes apparent when the sum of the incident and reflected waves is to be found.

K/θ curves

59. As the computation of the reflection coefficient is very tedious, curves of K_v , θ_v and K_h , θ_h , for the different states of polarization and for several different kinds of ground surface, are given in fig. 30 (Sheets 1 to 4). Once these are known the total field at a distance from an aerial may be found by the methods already given. For example, take a horizontal dipole operated at a frequency f , which is situated at a height h over ground for which σ and κ are known; and consider the total field at a point P at an angle φ to the horizon and at a distance r ($\gg h$) from the aerial. Let the aerial current be $I e^{j\omega t}$. Then, due to the aerial alone, we have at P a field

$$\gamma_A = \frac{60}{r} I e^{j\left(\omega t + \frac{\pi}{2} - \frac{2\pi}{\lambda} r\right)}.$$

Also, due to the wave reflected at the ground, a field

$$\gamma_B = \frac{60}{r} K_h I e^{j\left[\omega t + \frac{\pi}{2} - \frac{2\pi}{\lambda}(r + 2h \sin \varphi)\right]}.$$

Paying due regard to the conventional positive directions of γ_A and γ_B , therefore,

$$\begin{aligned} \gamma_T &= \frac{60}{r} I e^{j\left(\omega t + \frac{\pi}{2} - \frac{2\pi}{\lambda} r\right)} \left[1 - K_h e^{j\theta_h} e^{-j\frac{4\pi}{\lambda} h \sin \varphi}\right] \\ &= \gamma_A \left[1 - K_h / \theta_h - \frac{4\pi}{\lambda} h \sin \varphi\right]. \quad \dots \dots \dots (44) \end{aligned}$$

60. (i) By methods already explained, the amplitude of the total field is found to be

$$\hat{r}_T = \hat{r}_A \sqrt{1 + (K_h)^2 - 2K_h \cos\left(\theta_h - \frac{4\pi}{\lambda} h \sin \varphi\right)} \quad \dots \dots (45)$$

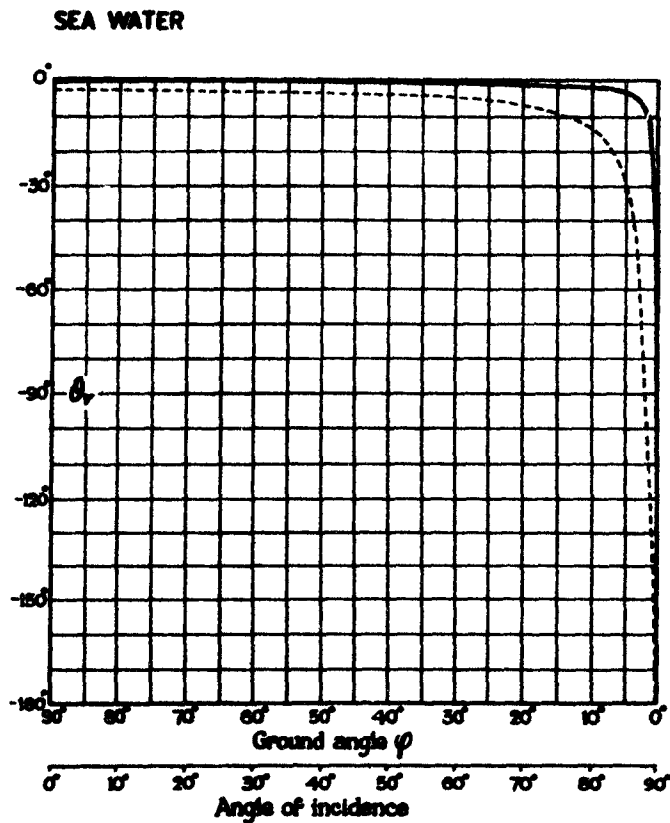
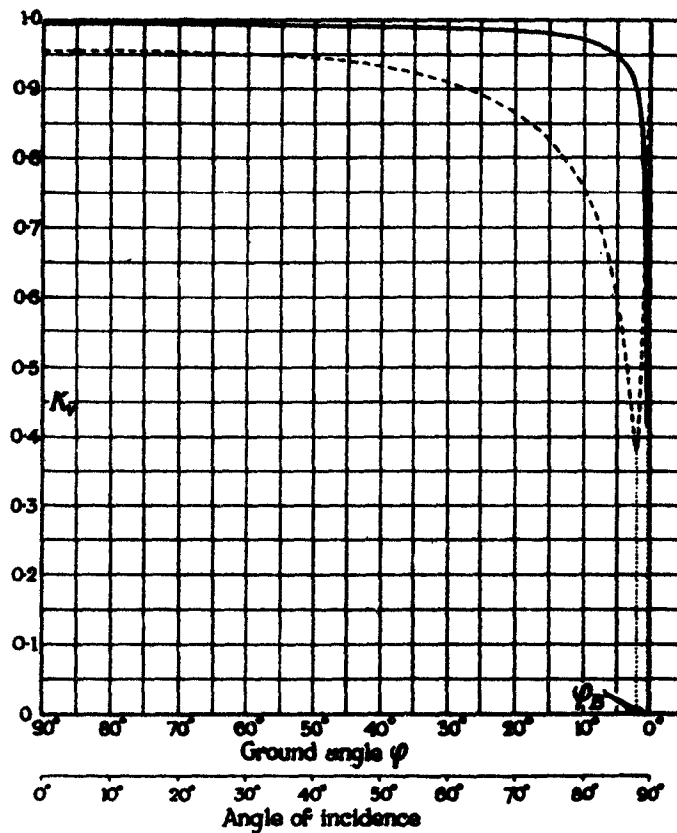
and we are not further concerned with its phase. The expression under the square root sign is therefore a factor by which the field strength $\hat{r}_A$, due to the aerial alone, must be multiplied, in order to allow for the effect of the earth. It is seen to be of the same form as that which takes into account the mutual impedance between an aerial and a reflector, but θ_h is a function of the angle φ instead of being constant for a given set of conditions.

(ii) If the above calculation is repeated for the case of a vertical dipole, with due regard to the sign convention, the amplitude of the total field is found to be

$$\hat{r}_T = \hat{r}_A \sqrt{1 + (K_v)^2 + 2K_v \cos\left(\theta_v - \frac{4\pi}{\lambda} h \sin \varphi\right)} \quad \dots \dots (46)$$

which is of a similar form to that obtained for horizontal polarization. The quantities under the square root signs are the Ground Reflection Factors and may be denoted by $q_v(\varphi)$ and $q_h(\varphi)$ respectively. Once these factors have been calculated, they may of course be applied to any array which in free space would radiate equally well above and below the equatorial plane, provided that h is taken as the height above earth of the electrical centre of the array. The R.M.S. field at any point having co-ordinates r , θ , φ , may therefore be written

$$r = \frac{60}{r} \cdot F \cdot f(\varphi) \cdot q(\varphi) \cdot I_A.$$



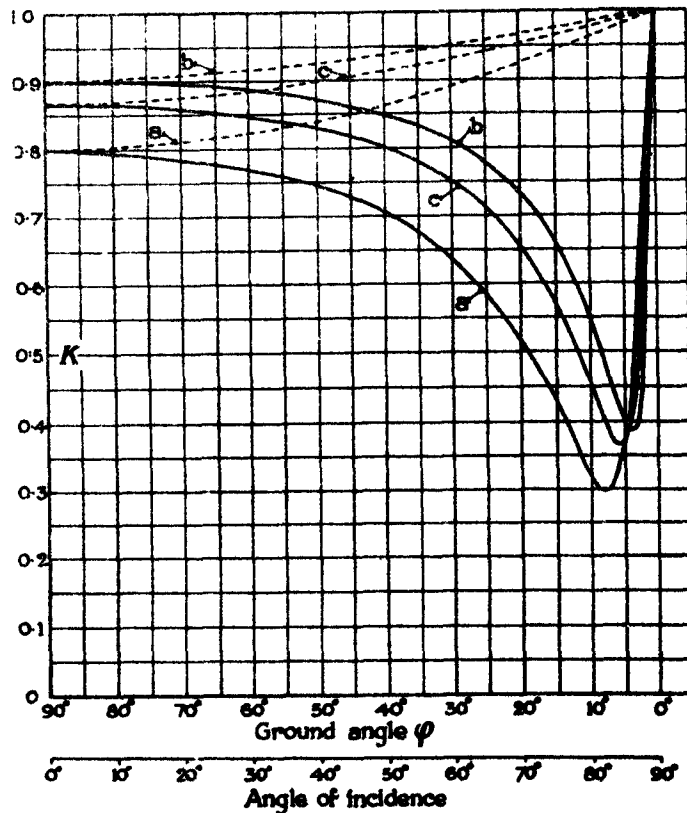
Γ IN PLANE OF PROPAGATION

$\cdots f = 10^8 \sim / \text{sec. } \sigma = 5 \times 10^{10} \text{ E.S.U. at } 0^\circ \text{C}$
 $\text{—} f = 10^8 \sim / \text{sec. } \sigma = 5 \times 10^{10} \text{ E.S.U. at } 25^\circ \text{C}$

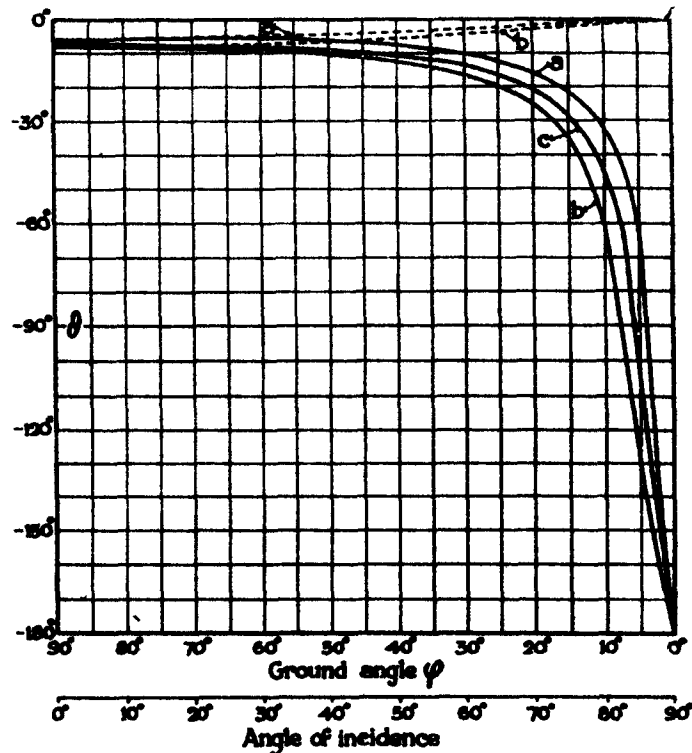
$\chi = 80$

REFLECTION COEFFICIENTS

FIG. 30
SHEET I
CHAP. IV



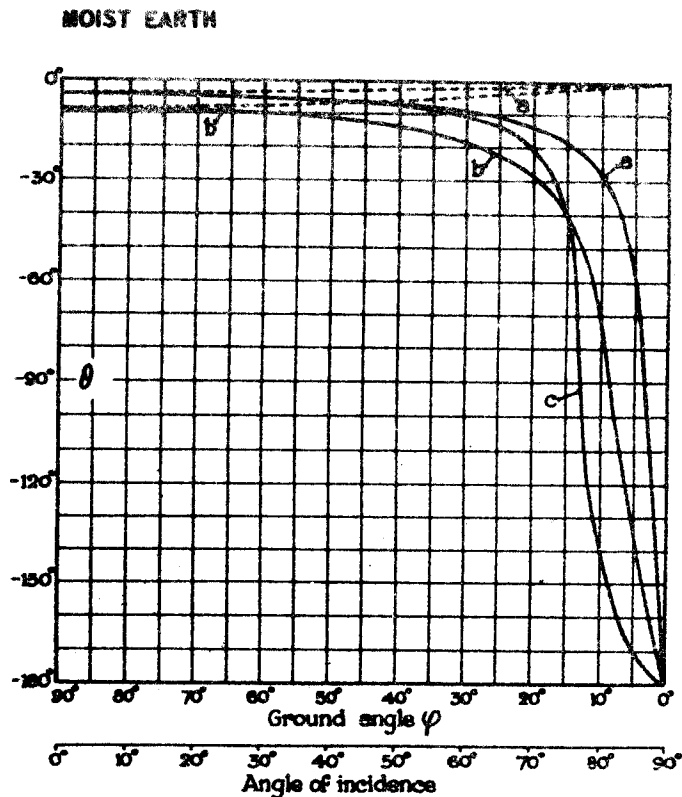
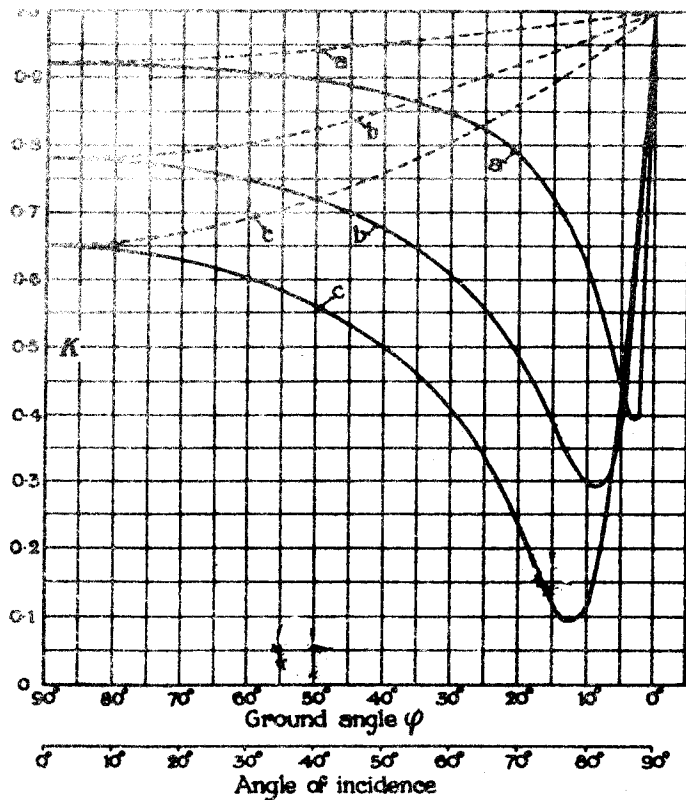
MOIST EARTH



— Γ IN PLANE OF PROPAGATION. --- Γ PERPENDICULAR TO PLANE OF PROPAGATION

- (a) $f = 10^6 \sim / \text{sec}$ $\sigma = 10^8 \text{ E.S.U.}$
 (b) $f = 10^7 \sim / \text{sec}$ $\sigma = 25 \times 10^8 \text{ E.S.U.}$
 (c) $f = 10^8 \sim / \text{sec}$ $\sigma = 55 \times 10^8 \text{ E.S.U.}$ } $\chi = 20$

REFLECTION COEFFICIENTS

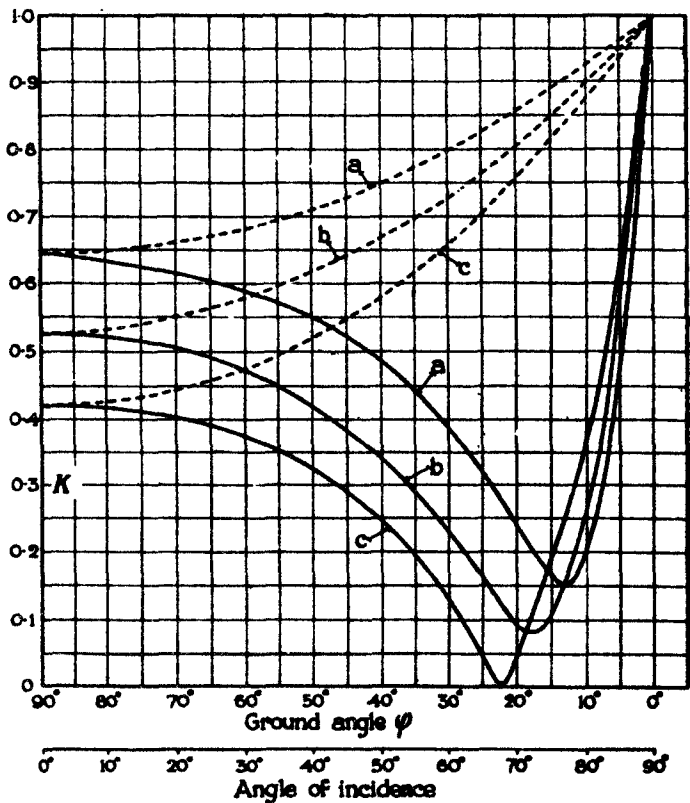


— Γ IN PLANE OF PROPAGATION

--- Γ PERPENDICULAR TO PLANE OF PROPAGATION

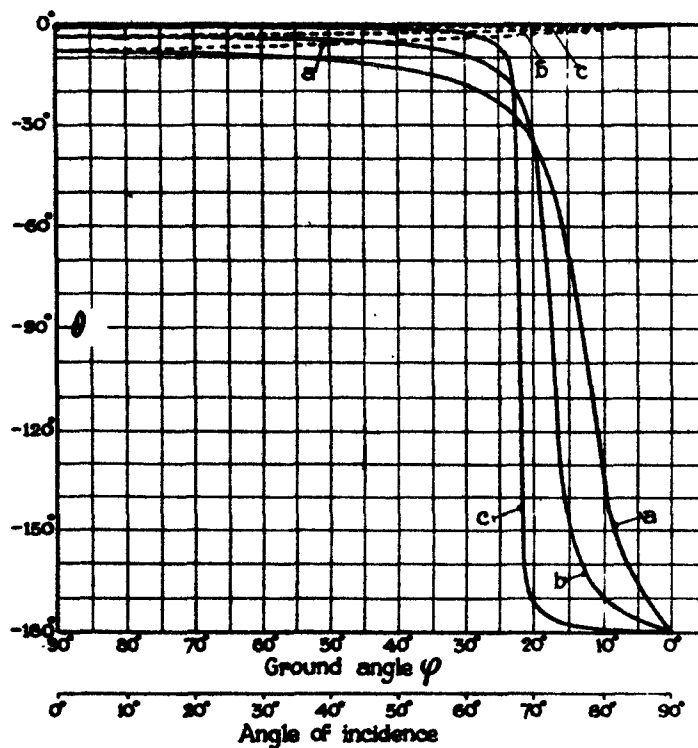
- | | | |
|-----------------------------------|---|---------------|
| (a) $f = 10^6 \sim / \text{sec.}$ | $\sigma = 1.5 \times 10^6 \text{ E.S.U.}$ | } $\chi = 20$ |
| (b) $f = 10^7 \sim / \text{sec.}$ | $\sigma = 2 \times 10^6 \text{ E.S.U.}$ | |
| (c) $f = 10^8 \sim / \text{sec.}$ | $\sigma = 4 \times 10^6 \text{ E.S.U.}$ | |

REFLECTION COEFFICIENTS



— Γ IN PLANE OF PROPAGATION

LOW CONDUCTIVITY SOILS



---- Γ PERPENDICULAR TO PLANE OF PROPAGATION

- | | | |
|----------------------------|---------------|------------------------------------|
| (a) Fine sands and gravels | $\kappa = 15$ | $\sigma = 0.5 \times 10^8$ E.S.U. |
| (b) Shingles and grits | $\kappa = 10$ | $\sigma = 0.1 \times 10^8$ E.S.U. |
| (c) Hard rocks | $\kappa = 6$ | $\sigma = 0.01 \times 10^8$ E.S.U. |

REFLECTION COEFFICIENTS

61. Referring now to the curves showing the variation of k_h with φ (figs. 30, Sheets 2 to 4) it is seen that for ground of moderately high conductivity, K_h may be of the order of 0.7 or more (for $\varphi = 90^\circ$), gradually increasing to unity when the radiation reaches the ground at a very small angle to the horizon. For very high values of σ , e.g. for sea water, K_h is rarely less than .96 (for $\varphi = 90^\circ$) and may therefore be taken as unity for practical purposes. For a surface of hard rock, on the other hand, K_h (for $\varphi = 90^\circ$) may be as low as 0.4 or even less. Given the values of κ , σ and f , it is not difficult to calculate the reflection coefficient for vertical incidence (i.e. $\varphi = 90^\circ$), as in the following example.

Example.—If $\sigma = 10^6$ E.S.U. $\kappa = 6$, $f = 10^7$ cycles per second, find the reflection coefficient for vertical incidence.

When $\varphi = 90^\circ$ $\cos \varphi = 0$, $\sin \varphi = 1$

$$k_h = \frac{\sqrt{\kappa - j \frac{2\sigma}{f}} - 1}{\sqrt{\kappa - j \frac{2\sigma}{f}} + 1}$$

$$\frac{2\sigma}{f} = 0.2$$

$$k_h = \frac{\sqrt{6 - j 0.2} - 1}{\sqrt{6 - j 0.2} + 1}$$

$$\sqrt{6 - j 0.2} = v - j\alpha$$

$$6 - j 0.2 = v^2 - 2jv\alpha - \alpha^2$$

$$\therefore v^2 - \alpha^2 = 6.$$

$$2v\alpha = 0.2$$

$$v^4 - 2v^2\alpha^2 + \alpha^4 = 36$$

$$4v^2\alpha^2 = 0.04$$

$$v^4 + 2v^2\alpha^2 + \alpha^4 = 36.04$$

$$v^2 + \alpha^2 = \sqrt{36.04}$$

$$= 6.0033$$

$$(v^2 + \alpha^2) + (v^2 - \alpha^2) = 12.0033$$

$$v^2 = 6.00165$$

$$v = \sqrt{6.00165}$$

$$= 2.45$$

and

$$\alpha^2 = 0.00165$$

$$\alpha = 0.0406$$

$$k_h = \frac{v - j\alpha - 1}{v - j\alpha + 1}$$

$$= \frac{2.45 - 1 - j 0.0406}{2.45 + 1 - j 0.0406}$$

$$= \frac{1.45 - j 0.0406}{3.45 - j 0.0406}$$

$$= \frac{1.45^2 + 0.0406^2}{3.45^2 + 0.0406^2}$$

$$K_h = \sqrt{\frac{1.45^2 + 0.0406^2}{3.45^2 + 0.0406^2}}$$

$$\doteq 0.42$$

CHAPTER XV.—PARAS. 62-63

When an approximate value for K_h ($\varphi = 90^\circ$) has been obtained, the approximate curve for other values of φ may be sketched in by noting that it closely resembles one quarter of the negative portion of a sine curve. The error in drawing the curve in this manner is greatest at about 30° , but even then is probably not greater than that occasioned by our imperfect knowledge of the electrical properties of the particular ground.

62. Turning now to the curves showing the value of K_v , it is at once evident that the phenomenon is more complicated than in the case of horizontal polarization. It is on this account that it is difficult to draw general conclusions as to the radiating properties of vertical aerials. The magnitude $K_{v(\max)}$ of the reflection coefficient for $\varphi = 90^\circ$ is the same as for horizontal polarization. This is obvious from physical considerations, for strictly, "vertical" and "horizontal" polarization have no significance for a wave perpendicularly incident. As the ground angle decreases, K_v also decreases and passes through a minimum value at some angle φ_b , afterwards increasing fairly rapidly, and reaching unity when $\varphi = 0$. The angle φ_b is known as the pseudo-Brewster angle from its relation to certain phenomena in optics. If the minimum value of K_v and the pseudo-Brewster angle φ_b are known for any particular kind of ground surface, the curve may be sketched in with sufficient accuracy for most purposes by observing the general trend of the calculated curves given. To facilitate this procedure, the curves shown in fig. 31 may be used.

63. The phase angles θ_v and θ_h are also plotted in fig. 30 for conditions corresponding to those for which K_v and K_h are given. Again it is obvious that for $\varphi = 90^\circ$, $\theta_v = \theta_h$ and is rarely more than a few degrees. In the case of horizontal polarization, the angle θ_h gradually decreases with an increase of φ and is zero when $\varphi = 0$. If its maximum value is known, the curve may be sketched in with fair accuracy by noting its general resemblance to a sine curve as in paragraph 62 above. As an example of the calculation we may find θ_h for the conditions previously discussed.

In paragraph 61 we found, for $\varphi = 90^\circ$

$$\begin{aligned} k_h &= \frac{1.45 - j 0.0406}{3.45 - j 0.0406} \\ &= \frac{(1.45 - j 0.0406)(3.45 + j 0.0406)}{3.45^2 + 0.0406^2} \\ &= \frac{5 + .00165 - j 0.14 + j 0.06}{11.9} \\ &\div \frac{5 - j 0.08}{11.9} \\ \therefore \theta_h &= -\tan^{-1} \frac{0.08}{5} \\ &= -0^\circ 55' \end{aligned}$$

The above example has been worked at some length, but the following "short-cut" should be

noted. When $\sqrt{\kappa - j \frac{2\sigma}{f}} = \nu - j\alpha$ has been evaluated

$$\begin{aligned} \frac{K_h/\theta_h}{(\varphi = 90^\circ)} &= \frac{\nu - 1 - j\alpha}{\nu + 1 - j\alpha} \\ &= \frac{\{(\nu - 1) - j\alpha\} \{(\nu + 1) + j\alpha\}}{(\nu + 1)^2 + \alpha^2} \\ &= \frac{\nu^2 + \alpha^2 - 1 - j 2\alpha}{(\nu + 1)^2 + \alpha^2} \end{aligned}$$

and when, as is generally the case in practice, $\nu^2 \gg \alpha^2$

$$K_h \div \frac{\nu - 1}{\nu + 1}$$

($\varphi = 90^\circ$)

$$\theta_h \div -\tan^{-1} \frac{2\alpha}{\nu^2 - 1}.$$

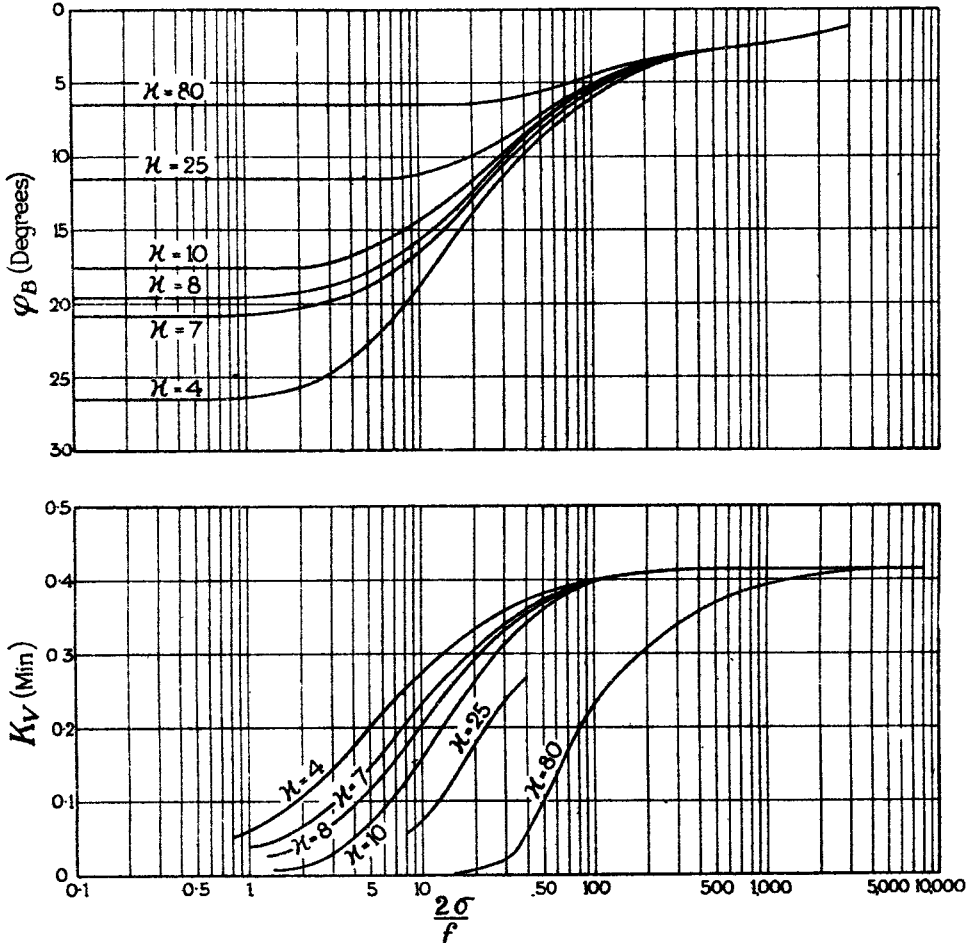


FIG. 31, CHAP. XV.—Pseudo-Brewster angle φ_B and corresponding reflection coefficient K_V (min.) for various kinds of ground.

64. The variation of θ_V with φ is very different from that of θ_h . Commencing with a small negative value equal to θ_h , when $\varphi = 90^\circ$, it is seen to increase in size very gradually until φ approaches the value φ_B , when a very rapid variation takes place; when $\varphi = \varphi_B$, $\theta_V = -90^\circ$, and for angles smaller than φ_B , θ_V continues to increase in size, becoming -180° when $\varphi = 0$. The minimum value of K_V and the corresponding angle φ_B are given in fig. 31 for various values of

κ and $\frac{2\sigma}{f}$ in order to facilitate the construction of approximate curves of K_V and θ_V .

CHAPTER XV.—PARAS. 65–66

65. (i) When the appropriate values of K and θ have been obtained and tabulated for any given conditions, the corresponding Reflection Factor can be plotted for various values of φ , and the resulting polar curve used as a correction factor for the free space diagram of any aerial or aerial array, taking the place of the Vertical Distribution Factor. As an example, the expression

$\sqrt{1 + K_v^2 + 2K_v \cos\left(\theta_v - \frac{4\pi}{\lambda} h \sin \varphi\right)}$ which is appropriate to vertically polarized radiation, has been plotted in thin solid line in fig. 32 for the following conditions, viz., $\kappa = 20$, $\sigma = 4 \times 10^8$, $f = 10^8$, $h = \frac{\lambda}{4}$. No great accuracy has been attempted as the intention is merely to indicate the

kind of curve to be expected. The curve shown in dotted line represents the Current Distribution Factor for some unspecified form of aerial. The curve shown in heavy line is the polar product of the two former curves; and gives the shape of the vertical polar diagram of the aerial or array.

Absolute values of R.M.S. field strength are of course obtained by multiplying by the factor $\frac{60}{r} F.I.$

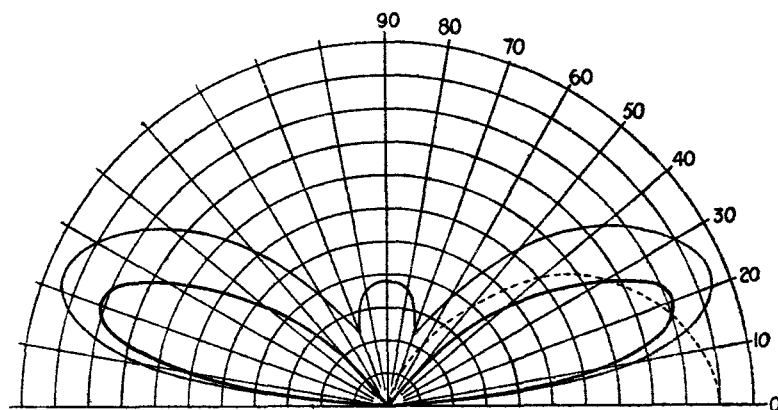


FIG. 32, CHAP. XV.—Vertical polar diagram of aerial over ground.
 $\sigma = 4 \times 10^8$, $\kappa = 20$, ($f = 10^8$, $h = \frac{\lambda}{4}$).

(ii) The foregoing theory assumes that at the boundary between the air and ground, the wave front is a plane surface. This incorrect assumption does not lead to significant error in the case of horizontal polarization, but with respect to vertically polarized waves, the field radiated along the surface of the earth is not absolutely zero as the simplified theory indicates. According to certain physicists, a vertical aerial at ground level gives rise to a surface wave additional to that derived from the simple radiation theory, but this view is not unreservedly accepted. Its protagonists agree that if this surface wave does exist, it suffers very heavy attenuation within a few wavelengths from the source, and need not be taken into account in long distance H/F and V. H/F communication.

TRANSMISSION LINES

Theory of transmission line

66. In Chapter VII reference is made to the propagation of electro-magnetic waves along a conductor such as a transmitting aerial. It is now necessary to enter somewhat more thoroughly into the theory of electro-magnetic waves on transmission lines such as the radio-frequency feeder lines used for supplying power from a transmitter to an aerial array, or from an aerial array to a radio receiver. The complete theory is also applicable to telephone and voice-frequency L/T lines. Before dealing with the mathematical theory, the physical aspect will be discussed.

67. Consider a transmission line consisting of a pair of parallel wires of high conductivity, perfectly insulated from and at a considerable height above the earth. Let these be connected to a battery by means of a reversing switch S as shown in fig. 33. A rapid reversal of the switch S is then equivalent to the application of an alternating E.M.F. having a perfectly flat-topped waveform. If the switch is closed at a given instant, so that the point A is at a positive potential with respect to the point B , an electric field will be set up between these points. The field does not however appear instantaneously at all points along the line, we may in fact consider the battery continuously to generate lines of electric force. Thus, if a single line of force appears between A and B when the switch is closed, and new lines are constantly being generated, the second line repels the first, causing the latter to travel along between the wires. As lines of electric force (unless closed upon themselves) must terminate upon electric charges, the movement of the electric lines implies the existence of moving electric charges, i.e. an electric current in the wires themselves. Thus, associated with the moving lines of electric force, we have a magnetic field consisting of a number of closed magnetic lines forming concentric circles round each conductor. The direction of the magnetic field relative to the direction of the electric field and the current is found by the first law of electro-dynamics.

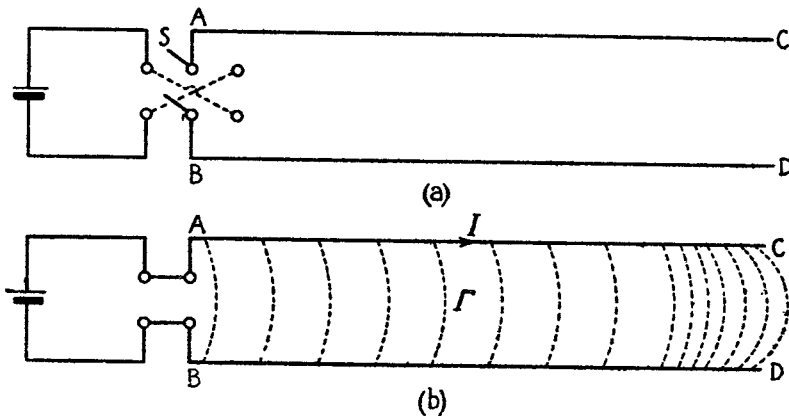


FIG. 33, CHAP. XV.—Electric field between parallel wires.

68. Now consider what happens when the travelling electric flux reaches the end of the line remote from the battery. If the wires are on open circuit, the lines of electric force can travel no further, and must tend to "pile up" at the points C D . In being brought to rest, however, they set up a magnetic field of opposite polarity to the original, and the growth of this field in turn recreates new lines of electric force. These lines now travel back towards the battery. This phenomenon may be summarized by the statement that on arrival at the open-circuited end of a transmission line, the electric field is reflected without change of phase, while the magnetic field is reflected with a phase change of 180° . At the moment of reversal, the magnetic field strength must fall to zero, and the electric field strength is doubled. If the remote end of the line is closed upon itself, forming what is called a short-circuited line, the reflection process is somewhat different. Instead of tending to pile up at the end, the electric lines must gradually collapse. In collapsing, however, they give rise to an additional magnetic field which travels on round the short-circuited end of the conductor. This magnetic field in turn recreates the electric field as before but with reverse polarity. At the exact instant at which the electric field is zero, the magnetic field strength is doubled. If the remote end of the line is connected to an impedance, partial reflection will occur, unless the terminal impedance has a particular nature and magnitude which will be dealt with later.

CHAPTER XV.—PARA. 69

General equations for line current and voltage

69. Although the foregoing physical aspect enables one to form a crude mental picture of the process of reflection it is necessary to enter somewhat more deeply into the effects of the line constants upon the mechanism of propagation. The line constants are first, the resistance, R (ohms per unit length), second, the inductance, L (henries per unit length), third, the capacitance, C (farads per unit length), between the two lines (or between line and earth in an "earth return" circuit), and fourth, the leakage conductance G (siemens, or mhos, per unit length). The resistance and inductance are measured per unit length of line, not per unit length of wire. It will easily be seen that in an element of line of length dx , the resistance will be $R.dx$, the inductance $L.dx$, the capacitance $C.dx$, and the leakage conductance $G.dx$ (fig. 34). Suppose, then, that an E.M.F. is applied to one end of the line, which will be called the input, or sending end, setting up at a

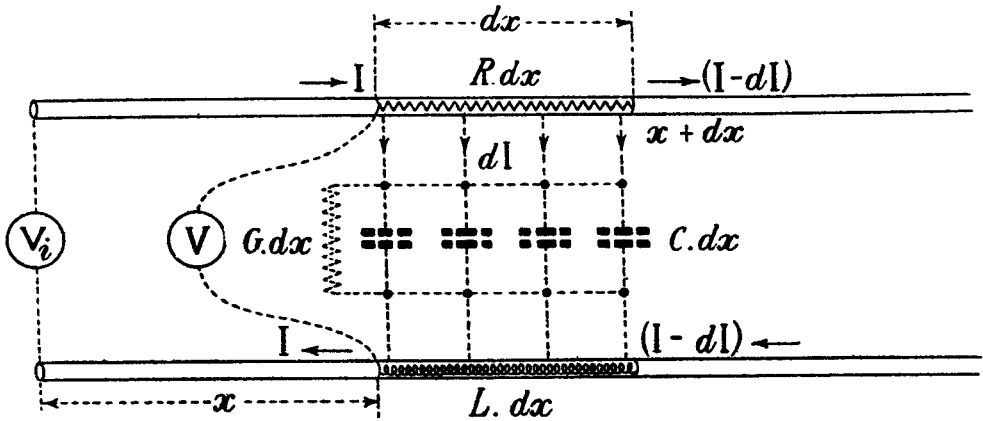


FIG. 34, CHAP. XV.—Notation used in transmission line theory.

point distant x centimetres along the line a P.D., V . If we now take an elementary length of line dx extending from x to $x + dx$, the P.D. between x and $x + dx$ will be $-dV$, where

$$-dV = (R.dx) I + (L.dx) \frac{dI}{dt}.$$

If the current in the line at the point x is I , it will be $I - dI$ at $x + dx$, owing to the element of current dI which flows in the capacitance $C.dx$ and leakage conductance $G.dx$. It is easily seen that

$$-dI = (G.dx) V + (C.dx) \frac{dV}{dt}.$$

Hence the rate of change of V and I , with respect to the distance from the sending end, is

$$-\frac{dV}{dx} = RI + L \frac{dI}{dt}$$

$$-\frac{dI}{dx} = GV + C \frac{dV}{dt}.$$

Solution for sinusoidal conditions

70. The above expressions are perfectly general, and subsequent work will be considerably simplified if the applied E.M.F. is considered to be sinusoidal. Under these conditions instead of the above equations we may write

$$-\frac{dV}{dx} = (R + j\omega L) I \quad \dots \dots \dots (1)$$

$$-\frac{dI}{dx} = (G + j\omega C) V \quad \dots \dots \dots (2)$$

because to a sinusoidal E.M.F., each unit length of line offers a vector impedance $R + j\omega L$, while shunted across each unit length we have a vector admittance $G + j\omega C$. Equations 1 and 2 are the fundamental basis of the theory of the transmission line. In developing the latter, we must first separate the variables V and I ; to do this differentiate equation 1 with respect to x :—

$$-\frac{d^2V}{dx^2} = (R + j\omega L) \frac{dI}{dx}$$

Substituting for $\frac{dI}{dx}$ from equation (2)

$$\begin{aligned} \frac{d^2V}{dx^2} &= (R + j\omega L) (G + j\omega C) V \\ &= P^2 V \end{aligned} \quad \dots \dots \dots (3)$$

In a similar manner we obtain

$$\frac{d^2I}{dx^2} = P^2 I \quad \dots \dots \dots (4)$$

71. These equations define $P = \sqrt{(R + j\omega L)(G + j\omega C)} = \alpha + j\beta$. It will be observed that P is complex and possesses the dimensions $\sqrt{\frac{\text{ohms}}{\text{length}} \times \frac{\text{siemens}}{\text{length}}} = \frac{1}{\text{length}}$, so that quantities like PI , Px , etc., are mere numbers. The complex quantity P is called the transfer constant of the line, and consists of a real part α called the attenuation constant, and an imaginary portion β called the wavelength constant, or latterly, the phase constant. Equations 3 and 4 are standard forms and the solutions are known to be

$$V = M_1 e^{-Px} + N_1 e^{Px} \quad \dots \dots \dots (5)$$

$$I = M_2 e^{-Px} + N_2 e^{Px} \quad \dots \dots \dots (6)$$

where M_1, M_2, N_1, N_2 , are quantities which depend upon the terminal conditions of the line, and are not entirely independent of each other. Since e^x is a mere number, M_1, M_2 , etc., must be vectors and are therefore printed in Clarendon. It will be shown that $M_1 = M_2 z_0, N_1 = -M_2 z_0$ where

$$z_0 = \sqrt{\frac{R + j\omega L}{G + j\omega C}}$$

Relation between M_1, M_2, N_1, N_2

72. If the values of V and I given in equations (5) and (6) are inserted in equation (1) we obtain

$$P(M_1 e^{-Px} - N_1 e^{Px}) = (R + j\omega L)(M_2 e^{-Px} + N_2 e^{Px}) \quad \dots \dots \dots (7a)$$

CHAPTER XV.—PARA. 73

and if inserted in equation (2)

$$P(\mathbf{M}_2 \varepsilon^{-Px} - \mathbf{N}_2 \varepsilon^{Px}) = (G + j\omega C)(\mathbf{M}_1 \varepsilon^{-Px} + \mathbf{N}_1 \varepsilon^{Px}). \quad \dots \dots \dots (7b)$$

On multiplying across by $\frac{P}{G + j\omega C}$

$$\begin{aligned} P(\mathbf{M}_1 \varepsilon^{-Px} + \mathbf{N}_1 \varepsilon^{Px}) &= \frac{P^2}{G + j\omega C}(\mathbf{M}_2 \varepsilon^{-Px} - \mathbf{N}_2 \varepsilon^{Px}) \\ &= (R + j\omega L)(\mathbf{M}_2 \varepsilon^{-Px} - \mathbf{N}_2 \varepsilon^{Px}). \quad \dots \dots \dots (8) \end{aligned}$$

Adding equations (7a) and (8)

$$\begin{aligned} 2P\mathbf{M}_1 \varepsilon^{-Px} &= (R + j\omega L) \times 2\mathbf{M}_2 \varepsilon^{-Px} \\ \therefore \mathbf{M}_1 &= \mathbf{M}_2 \frac{R + j\omega L}{P} \\ &= \mathbf{M}_2 \sqrt{\frac{R + j\omega L}{G + j\omega C}} \\ &= \mathbf{M}_2 \mathbf{z}_0. \end{aligned}$$

Subtracting (8) from (7a)

$$\begin{aligned} -2P\mathbf{N}_1 \varepsilon^{Px} &= + (R + j\omega L) \times 2\mathbf{N}_2 \varepsilon^{Px} \\ \therefore \mathbf{N}_1 &= -\mathbf{N}_2 \frac{R + j\omega L}{P} \\ &= -\mathbf{N}_2 \mathbf{z}_0. \end{aligned}$$

The quantity $\mathbf{z}_0 = \sqrt{\frac{R + j\omega L}{G + j\omega C}}$ is of great importance; it is termed the characteristic impedance or surge impedance of the line.

Introduction of hyperbolic functions

73. Equations (5) and (6) may now be written

$$\mathbf{V} = \mathbf{M}_1 \varepsilon^{-Px} + \mathbf{N}_1 \varepsilon^{Px} \quad \dots \dots \dots (9)$$

$$\mathbf{I} = \frac{\mathbf{M}_1}{\mathbf{z}_0} \varepsilon^{-Px} - \frac{\mathbf{N}_1}{\mathbf{z}_0} \varepsilon^{Px} \quad \dots \dots \dots (10)$$

It is now convenient to introduce hyperbolic functions, writing

$$\begin{aligned} \cosh Px + \sinh Px &= \varepsilon^{Px} \\ \cosh Px - \sinh Px &= \varepsilon^{-Px} \end{aligned}$$

so that (9) and (10) become

$$\begin{aligned} \mathbf{V} &= \mathbf{M}_1 (\cosh Px - \sinh Px) + \mathbf{N}_1 (\cosh Px + \sinh Px) \\ &= (\mathbf{M}_1 + \mathbf{N}_1) \cosh Px - (\mathbf{M}_1 - \mathbf{N}_1) \sinh Px, \quad \dots \dots \dots (11) \end{aligned}$$

$$\mathbf{I} = \frac{1}{\mathbf{z}_0} [(\mathbf{M}_1 - \mathbf{N}_1) \cosh Px - (\mathbf{M}_1 + \mathbf{N}_1) \sinh Px] \quad \dots \dots (12)$$

Note that we have reduced the number of quantities depending upon the terminal conditions to two. Provided $\mathbf{M}_1$ and $\mathbf{N}_1$ can be determined we are able to obtain complete information regarding the current and voltage distribution in the line.

Equations for infinite line

74. The simplest problem to consider, and one of great importance because it brings out the physical signification of $\mathbf{z}_0$ and P , is a line of infinite length to which a known voltage $\mathbf{V}_1$ (i for "input") is applied to the sending end. At the point $x = 0$ we have $\mathbf{V} = \mathbf{V}_1$; inserting known quantities in equation (11)

$$\mathbf{V}_1 = (\mathbf{M}_1 + \mathbf{N}_1) \cosh 0 - (\mathbf{M}_1 - \mathbf{N}_1) \sinh 0$$

$$(\cosh 0 = 1, \sinh 0 = 0)$$

$$\mathbf{V}_1 = \mathbf{M}_1 + \mathbf{N}_1$$

On the other hand, it is obvious from physical reasons that as we go further from the sending end both $\mathbf{V}$ and $\mathbf{I}$ become smaller, and ultimately when $x \rightarrow \infty$, $\mathbf{V} \rightarrow 0$, $\mathbf{I} \rightarrow 0$. But when $x \rightarrow \infty$, both $\cosh Px$ and $\sinh Px$ approach the value $\frac{1}{2} e^{Px}$, and therefore

$$\mathbf{V}_{(x \rightarrow \infty)} = [(\mathbf{M}_1 + \mathbf{N}_1) - (\mathbf{M}_1 - \mathbf{N}_1)] \frac{e^{Px}}{2} = 0.$$

Now $\frac{1}{2} e^{Px}$ is not equal to zero, therefore

$$\mathbf{M}_1 + \mathbf{N}_1 - (\mathbf{M}_1 - \mathbf{N}_1) = 0$$

$$\therefore \mathbf{N}_1 = 0.$$

But $\mathbf{M}_1 + \mathbf{N}_1 = \mathbf{V}_1$

$$\therefore \mathbf{V}_1 = \mathbf{M}_1.$$

Inserting in (11)

$$\begin{aligned} \mathbf{V} &= \mathbf{M}_1 \cosh Px - \mathbf{M}_1 \sinh Px \\ &= \mathbf{V}_1 (\cosh Px - \sinh Px) \\ &= \mathbf{V}_1 e^{-Px}, \end{aligned} \quad \dots \dots \dots (13)$$

and in (12),

$$\begin{aligned} \mathbf{I} &= \frac{\mathbf{V}_1}{\mathbf{z}_0} (\cosh Px - \sinh Px) \\ &= \frac{\mathbf{V}_1}{\mathbf{z}_0} e^{-Px}. \end{aligned} \quad \dots \dots \dots (14)$$

Magnitudes of α and β

75. As already stated $P = \sqrt{(R + j\omega L)(G + j\omega C)} = \alpha + j\beta$. It follows that

$$RG + j\omega CR + j\omega LG - \omega^2 LC = \alpha^2 + 2j\alpha\beta - \beta^2$$

$$\alpha^2 - \beta^2 = RG - \omega^2 LC \quad \dots \dots \dots (15a)$$

$$2\alpha\beta = \omega(CR + LG)$$

$$(RG - \omega^2 LC)^2 = \alpha^4 - 2\alpha^2\beta^2 + \beta^4$$

$$\omega^2(CR + LG)^2 = 4\alpha^2\beta^2$$

$$\alpha^2 + \beta^2 = \sqrt{(RG - \omega^2 LC)^2 + \omega^2(CR + LG)^2} \quad \dots \dots \dots (15b)$$

CHAPTER XV.—PARAS. 76-77

From equations 15a and 15b we obtain

$$\alpha = \sqrt{\frac{1}{2} \left\{ \sqrt{(R^2 + \omega^2 L^2)(G^2 + \omega^2 C^2)} + (GR - \omega^2 LC) \right\}} \quad \dots (15c)$$

$$\beta = \sqrt{\frac{1}{2} \left\{ \sqrt{(R^2 + \omega^2 L^2)(G^2 + \omega^2 C^2)} - (GR - \omega^2 LC) \right\}}. \quad \dots (15d)$$

Physical significance of α and β

76. Equation 13 may be written

$$\begin{aligned} \mathbf{V} &= \mathbf{V}_1 e^{-(\alpha + j\beta)x} \\ &= \mathbf{V}_1 e^{-\alpha x} e^{-j\beta x} \\ &= (\mathbf{V}_1 e^{-\alpha x}) (\cos \beta x - j \sin \beta x). \end{aligned}$$

The portion within the first pair of brackets may be called the amplitude factor. It indicates that the amplitude $\mathcal{V}$ of the voltage at x is equal to the amplitude $\mathcal{V}_1$ of the input voltage divided by $e^{\alpha x}$. The factor within the second pair of brackets is a vector operator of unit magnitude. Its presence signifies that the phase of $\mathbf{V}$ lags behind that of $\mathbf{V}_1$ by an angle βx . Thus if

$$\begin{aligned} v_1 &= \mathcal{V}_1 \cos(\omega t + \varphi) \\ v_x &= \frac{\mathcal{V}_1}{e^{\alpha x}} \cos(\omega t + \varphi - \beta x). \end{aligned}$$

In a very long or infinite line, therefore, the voltage amplitude at a distance x from the input end decreases exponentially by the factor $e^{-\alpha x}$, while the phase angle lags behind the phase at the sending end by an angle βx . At a distance such that $\beta x = 2\pi$ the line voltage is in phase with the supply voltage. Similarly if $\beta x = 4\pi, 6\pi$, etc., in fact, the line voltage is in phase with that at the sending ends at all points where $\beta x = 2\pi n$, and n is an integer.

Physical significance of z_0 .

77. We may now consider the current in an infinite line. At the sending end, where $x = 0$, let it be $\mathbf{I}_1$. From equation (14)

$$\mathbf{I}_1 = \frac{\mathbf{V}_1}{z_0} e^0 = \frac{\mathbf{V}_1}{z_0} \quad \dots \dots \dots (16a)$$

Hence z_0 is the quotient of voltage and current at the input end of an infinite line, and is therefore the input impedance of such a line. It also follows that equation 14 may be written

$$\mathbf{I}_x = \mathbf{I}_1 e^{-Px} \quad \dots \dots \dots (17)$$

This equation is of exactly the same form as equation 13 and may be interpreted in the same way, i.e. in passing along the line the current is attenuated and its phase delayed just as in the case of the voltage. Further, at any point in the line

$$\frac{\mathbf{V}_x}{\mathbf{I}_x} = z_0 \quad \dots \dots \dots (16b)$$

thus z_0 is the ratio of voltage to current at any point in the line.

Short-circuited lines

78. We have now shown the physical meanings of the quantities α , β and z_0 , and may apply these to more practical cases, e.g. a line of finite length terminated by an impedance of some kind. Consider a length of line l which is short-circuited at the output or receiving end. Let a voltage V_1 be applied at the point $x = 0$.

Then the following data are known :—

$$\text{at } x = 0, V = V_0 = V_1$$

$$\text{at } x = l, V = V_1 = 0.$$

Inserting these conditions in equation (11)

$$V_1 = M_1 + N_1$$

$$V_1 = 0 = (M_1 + N_1) \cosh Pl - (M_1 - N_1) \sinh Pl$$

$$= V_1 \cosh Pl - (M_1 - N_1) \sinh Pl$$

$$\therefore M_1 - N_1 = \frac{V_1 \cosh Pl}{\sinh Pl}$$

$$= V_1 \coth Pl \dots \dots \dots (18)$$

so that equation 11 may be written

$$V = V_1 (\cosh Px - \coth Pl \sinh Px)$$

$$= V_1 \frac{\sinh P(l-x)}{\sinh Pl}$$

and equation 12 becomes

$$I = \frac{V_1}{z_0} (\coth Pl \cosh Px - \sinh Px)$$

$$= \frac{V_1}{z_0} \frac{\cosh P(l-x)}{\sinh Pl}.$$

Now at the input end, $x = 0$. The current entering the line is therefore

$$I_1 = \frac{V_1}{z_0} \frac{\cosh Pl}{\sinh Pl}$$

$$= \frac{V_1}{z_0 \tanh Pl} \dots \dots \dots (19)$$

Thus we have a most important result, namely that the input impedance of a length l of line short-circuited at the end remote from the input terminals, is $z_0 \tanh Pl$.

Open lines

79. Another case of interest is that of a line of length l , with the output end on open circuit. At the input end $x = 0, V = V_1$, while at the output end $x = l, I = 0$.

From equation (11)

$$V_1 = M_1 + N_1$$

$$0 = \frac{1}{z_0} ((M_1 - N_1) \cosh Pl - (M_1 + N_1) \sinh Pl)$$

$$M_1 - N_1 = (M_1 + N_1) \frac{\sinh Pl}{\cosh Pl}$$

$$= V_1 \tanh Pl \dots \dots \dots (20)$$

CHAPTER XV.—PARA. 80

and the line equations become

$$V_x = V_i \cosh Px - V_i \tanh Pl \sinh Px$$

$$I_x = \frac{V_i}{z_0} \{ \tanh Pl \cosh Px - \sinh Px \}$$

or

$$V_x = V_i \frac{\cosh P(l-x)}{\cosh Pl}$$

$$I_x = \frac{V_i}{z_0} \frac{\sinh P(l-x)}{\cosh Pl}$$

The input current is obtained by putting $x = 0$, hence

$$\begin{aligned} I_i &= \frac{V_i}{z_0} \frac{\sinh Pl}{\cosh Pl} \\ &= \frac{V_i}{z_0 \coth Pl} \quad \dots \dots \dots (21) \end{aligned}$$

Thus the input impedance of a length l of open-circuited line is $z_0 \coth Pl$.

Line terminated by finite impedance

80. Having cleared the air by these preliminary investigations we arrive at the most important practical case, namely, a line of finite length l , terminated by a finite impedance of z_r ohms, its nature being unspecified. As before it is known that $V = V_i$ at $x = 0$. At the other end, where $x = l$, the current and voltage will be denoted by I_r and V_r (r for "receiving"). Then I_r is the current through z_r due to the P.D. V_r , and

$$I_r = \frac{V_r}{z_r}.$$

Putting $V = V_i$ when $x = 0$,

$$M_1 + N_1 = V_i$$

Since $V_r = z_r I_r$ it follows that when $x = l$, equations (11) and (12) become

$$\begin{aligned} V_i \cosh Pl - (M_1 - N_1) \sinh Pl &= \frac{z_r}{z_0} \left\{ (M_1 - N_1) \cosh Pl - V_i \sinh Pl \right\} \\ V_i \left\{ \cosh Pl + \frac{z_r}{z_0} \sinh Pl \right\} &= (M_1 - N_1) \left\{ \sinh Pl + \frac{z_r}{z_0} \cosh Pl \right\} \\ M_1 - N_1 &= \frac{V_i \left\{ \cosh Pl + \frac{z_r}{z_0} \sinh Pl \right\}}{\sinh Pl + \frac{z_r}{z_0} \cosh Pl} \quad \dots \dots (22) \end{aligned}$$

hence for these conditions, equations (13) and (14) become

$$V = V_i \frac{z_r \cosh P(l-x) + z_0 \sinh P(l-x)}{z_r \cosh Pl + z_0 \sinh Pl} \quad \dots \dots (23)$$

and

$$I = \frac{V_i}{z_0} \frac{z_r \sinh P(l-x) + z_0 \cosh P(l-x)}{z_r \cosh Pl + z_0 \sinh Pl} \quad \dots \dots (24)$$

Equations (23) and (24) give the voltage and current at any distance x along the line. For many purposes we require to know only those at the input and output ends respectively. Putting $x = l$ in (23) and (24),

$$V_r = V_i \frac{z_r}{z_r \cosh Pl + z_0 \sinh Pl} \quad \dots \quad (25)$$

$$I_r = V_i \frac{1}{z_r \cosh Pl + z_0 \sinh Pl} \quad \dots \quad (26)$$

while putting $x = 0$ in (24) gives

$$I_i = \frac{V_i}{z_0} \frac{z_r \sinh Pl + z_0 \cosh Pl}{z_r \cosh Pl + z_0 \sinh Pl} \quad \dots \quad (27a)$$

The input impedance is therefore

$$z_i = z_0 \frac{z_r \cosh Pl + z_0 \sinh Pl}{z_r \sinh Pl + z_0 \cosh Pl} \quad \dots \quad (27b)$$

Correctly terminated line

81. A very important case in practice is that which occurs when the terminating load z_r is equal to the surge impedance z_0 of the line. When this is so, equation (24) becomes

$$\begin{aligned} I_i &= \frac{V_i}{z_0} \frac{z_0 (\sinh Pl + \cosh Pl)}{z_0 (\sinh Pl + \cosh Pl)} \\ &= \frac{V_i}{z_0} \quad \dots \quad (28a) \end{aligned}$$

while

$$\begin{aligned} I_r &= \frac{V_i}{z_0 (\cosh Pl + \sinh Pl)} \\ &= \frac{V_i}{z_0 \epsilon^{Pl}} \\ &= \frac{V_i}{z_0} \epsilon^{-Pl} \quad \dots \quad (28b) \end{aligned}$$

This is exactly the same expression as was found in paragraph 77 for the current I_r at a distance x from the input end of an infinitely long line. It follows then that if a line of finite length is terminated by an impedance equal to its surge impedance, all the energy reaching the output terminals of the line passes into the load impedance, which is usually the desired object. When the operating conditions are such that $z_r = z_0$ the line is said to be correctly terminated. When incorrectly terminated the whole of the received energy does not pass the output terminals, a portion being reflected back towards the input end. The importance of avoiding reflection in a transmission line may perhaps be emphasized by comparing it with an aerial. With a few exceptions, aërials are built up of conductors with free ends, so that reflection occurs, and the

length (including the image in certain cases) is made electrically equal to a multiple of $\frac{\lambda}{2}$ so that

stationary waves are set up in the aerial. By this means we obtain syn-phased currents over each half-wavelength of wire (approximately) as explained in the early paragraphs of this chapter. The energy supplied to the aerial is then partly radiated and partly degraded into heat. In a transmission line, however, the object is to convey as much energy as possible from one point (the transmitter) to another point (the load impedance), avoiding all unnecessary dissipation en route. For this conveyance to be highly efficient, then, the load impedance must be equal to the surge impedance of the line.

CHAPTER XV.—PARAS. 82-84

82. Returning now to the expression for the current in the load impedance in the case of a correctly terminated line, i.e.

$$I_r = \frac{V_i}{Z_0} e^{-Pl}$$

putting

$$P = \alpha + j\beta$$

$$I_r = \frac{V_i}{Z_0} e^{-\alpha l} e^{-j\beta l}.$$

We have already seen that the magnitude of the attenuation constant α depends upon the leakage conductance G and resistance R , per unit length. If G , R and l are very small $e^{-\alpha l}$ is very nearly unity and the received current becomes

$$I_r = \frac{V_i}{Z_0} e^{-j\beta l}. \quad \dots \dots \dots (29)$$

At radio frequencies, α is given by the following approximate formula which is derived from equation (15c);

$$\alpha = \frac{R}{2Z_0} + \frac{GZ_0}{2},$$

and is always a very small quantity. Suppose the line to have a resistance of 20 ohms per mile and an insulation resistance of 5 megohms per mile. One mile is roughly 1.6×10^5 centimetres, i.e. $R = 1.2 \times 10^{-4}$ ohms, $\frac{1}{G} = 5 \times 10^6 \times 1.6 \times 10^5$ ohms per centimetre, and $G = 1.25 \times 10^{-12}$ siemens per centimetre. If the surge impedance of the line is 500 ohms, the attenuation constant is

$$\begin{aligned} \alpha &= \frac{1.2}{2 \times 10^4 \times 500} + \frac{1.25 \times 500}{10^{12} \times 2} \\ &= 1.203 \times 10^{-7}. \end{aligned}$$

Radio-frequency feeders

83. In connecting an aerial array to its transmitting or receiving equipment, it is necessary to utilize a transmission line consisting of either a twin wire line or a concentric line. In either instance the length rarely exceeds a few hundred feet, and the line may be designed to have a very low attenuation constant. The theory may then be considerably simplified by assuming the attenuation to be negligible, i.e. that the line itself has negligible resistance and perfect insulation so that in the equation $P = \sqrt{(R + j\omega L)(G + j\omega C)}$, $R = 0$ and $G = 0$. Then $P = \alpha + j\beta = \sqrt{j\omega L \times j\omega C} = j\omega\sqrt{LC}$, and therefore $\alpha = 0$, $\beta = \omega\sqrt{LC}$. Similarly the equation $Z_0 = \sqrt{\frac{R + j\omega L}{G + j\omega C}}$ becomes $Z_0 = \sqrt{\frac{L}{C}}$. In these circumstances Z_0 is not complex and therefore possesses no reactive component, i.e. the surge impedance is purely resistive, and may be denoted by Z_0 . When its non-reactive nature is to be particularly stressed it will be denoted by R_0 .

Twin wire feeders

84. The inductance of a pair of parallel wires of radius r , separated by a distance D , is

$$L = \frac{9.2104}{10^9} \log_{10} \frac{D}{r} \text{ henries per centimetre,}$$

and the capacitance (assuming the dielectric to be air) is

$$C = \frac{1 \cdot 208}{10^{12} \log_{10} \frac{D}{r}} \text{ farads per centimetre,}$$

so that

$$\begin{aligned} Z_0 &= \sqrt{\frac{L}{C}} = \sqrt{\frac{9 \cdot 2104}{10^9} \log_{10} \frac{D}{r} \times \frac{10^{12}}{1 \cdot 208} \log_{10} \frac{D}{r}} \\ &= 276 \log_{10} \frac{D}{r} \text{ ohms.} \end{aligned}$$

The phase constant of the line is easily found :—

$$\begin{aligned} \beta &= \omega \sqrt{LC} \\ &= 2\pi f \sqrt{\frac{9 \cdot 2104}{10^9} \times \frac{1 \cdot 208}{10^{12}}} \\ &= \frac{2\pi f}{3 \times 10^{10}} \\ &= \frac{2\pi f}{c} \end{aligned}$$

where c is the natural constant equal to the velocity of electro-magnetic waves in free space.

Since $\frac{f}{c} = \frac{1}{\lambda}$, where λ is the wavelength in free space

$$\beta = \frac{2\pi}{\lambda}.$$

Concentric feeders

85. A concentric feeder consists of an outer tubular conductor containing an inner conductor which may be either solid or tubular. If D is the internal diameter of the outer tube and d the external diameter of the inner conductor, its inductance and capacitance are given by the formulae

$$L = \frac{4 \cdot 605}{10^9} \log_{10} \frac{D}{d} \text{ henries per centimetre}$$

$$C = \frac{2 \cdot 416}{10^{12} \log_{10} \frac{D}{d}} \text{ farads per centimetre}$$

and, therefore,

$$Z_0 = 138 \log_{10} \frac{D}{d} \text{ ohms.}$$

It is easily shown that, as for the twin wire feeder,

$$\beta = \frac{2\pi}{\lambda}.$$

As stated in Chapter VII, where twin wire feeders are used, it is usual to arrange, if possible, that the surge impedance is 600 ohms. This implies that the ratio $\frac{D}{r} = 150$. For example, 18 s.w.g. wire has a diameter of .048 inch, and gives a surge impedance of 600 ohms if spaced 3.6 inches

CHAPTER XV.—PARA. 86

apart. The surge impedance of a concentric feeder is usually about 60 to 100 ohms, and the copper losses are a minimum when $\frac{D}{d} = 3.6$, i.e. when the surge impedance is 75 ohms. For ratios smaller than 2 the copper losses are very heavy, but they are not seriously increased by an increase of $\frac{D}{d}$ up to about 8.

Properties of lines of various lengths

86. (i) The input impedance of a length l of line, short-circuited at the output end, is $z_1 = z_0 \tanh Pl$. If the attenuation is negligible, $P = j\beta = j\frac{2\pi}{\lambda}$,

$$z_1 = z_0 \tanh j \frac{2\pi}{\lambda} l.$$

Since, however, z_0 is a purely ohmic resistance of $\sqrt{\frac{L}{C}}$ ohms,

$$\begin{aligned} z_1 &= \sqrt{\frac{L}{C}} \tanh j \frac{2\pi}{\lambda} l \\ &= j \sqrt{\frac{L}{C}} \tan \frac{2\pi}{\lambda} l, \dots \dots \dots (30) \end{aligned}$$

i.e. z_1 is purely reactive and may be either positive or negative. The graph of the magnitude Z_1 of the input reactance, against the length l , is plotted in fig. 35. It is seen that at the point $l = 0$, $Z_1 = 0$. As l increases Z_1 assumes positive values, e.g. at $l = \frac{\lambda}{8}$, $Z_1 = Z_0$, and increases until at $l = \frac{\lambda}{4}$, Z_1 becomes infinite. In the range $l = 0$ to $l = \frac{\lambda}{4}$ then, the line behaves as an inductance, the value of which may lie anywhere between zero and infinity. Consequently, a

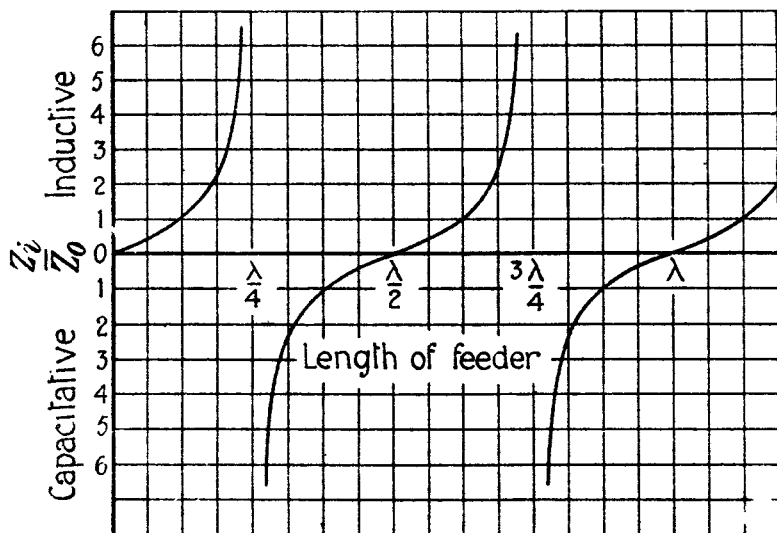


FIG. 35. CHAP. XV.—Reactance of short-circuited line.

length of line may be determined which will act as an inductance of any desired value for a given frequency. Suppose we desire a line to have an inductance L' ,

$$\begin{aligned}j\omega L' &= jZ_0 \tan \frac{2\pi}{\lambda} l \\ \tan \frac{2\pi}{\lambda} l &= \frac{\omega L'}{Z_0} \\ \frac{2\pi}{\lambda} l &= \tan^{-1} \frac{\omega L'}{Z_0} \\ l &= \frac{\lambda}{2\pi} \tan^{-1} \frac{\omega L'}{Z_0}.\end{aligned}$$

Example

Calculate the length of 600 ohms line which will act as an inductance of $5\mu H$ at a frequency of 5 M/cs.

$$\begin{aligned}\lambda &= \frac{3 \times 10^8}{5 \times 10^6} = 60 \text{ metres} \\ l &= \frac{60}{2\pi} \tan^{-1} \frac{2\pi \times 5 \times 10^6 \times 5 \times 10^{-6}}{600} \\ &= 9.56 \tan^{-1} \frac{\pi}{12} \\ \tan^{-1} \frac{\pi}{12} &= 14^\circ 41' \text{ or } 0.25 \text{ radians} \\ l &= 9.56 \times 0.25 \\ &= 2.4 \text{ metres.}\end{aligned}$$

In the range $\frac{\lambda}{4}$ to $\frac{\lambda}{2}$, $\tan \frac{2\pi}{\lambda} l$ is negative and the reactance of the short-circuited line is capacitive, varying from infinity to zero. By a suitable choice of l , its reactance may be of any value whatever. If it is required to obtain a line of capacitance C' farads,

$$\begin{aligned}\frac{1}{j\omega C} &= jZ_0 \tan \frac{2\pi}{\lambda} l \\ \tan \frac{2\pi}{\lambda} l &= -\frac{1}{\omega C' Z_0} \\ l &= \frac{\lambda}{2\pi} \tan^{-1} \left(-\frac{1}{\omega C' Z_0} \right)\end{aligned}$$

Over the range $\frac{\lambda}{2}$ to λ the curve repeats the values in the range 0 to $\frac{\lambda}{2}$ and so on.

(ii) The input impedance of a length of line having its output end free, is $Z_1 = Z_0 \coth Pl$. For negligible attenuation this becomes

$$Z_1 = -jZ_0 \cot \frac{2\pi}{\lambda} l. \quad \dots \dots \dots (31)$$

CHAPTER XV.—PARA. 87

As before, then, the impedance is purely reactive ; Z_i is plotted against l in fig. 36. By using a length of line less than $\frac{\lambda}{4}$ we may obtain a negative (capacitive) reactance of any value between $-\infty$ and 0 while lengths between $\frac{\lambda}{4}$ and $\frac{\lambda}{2}$ behave inductively.

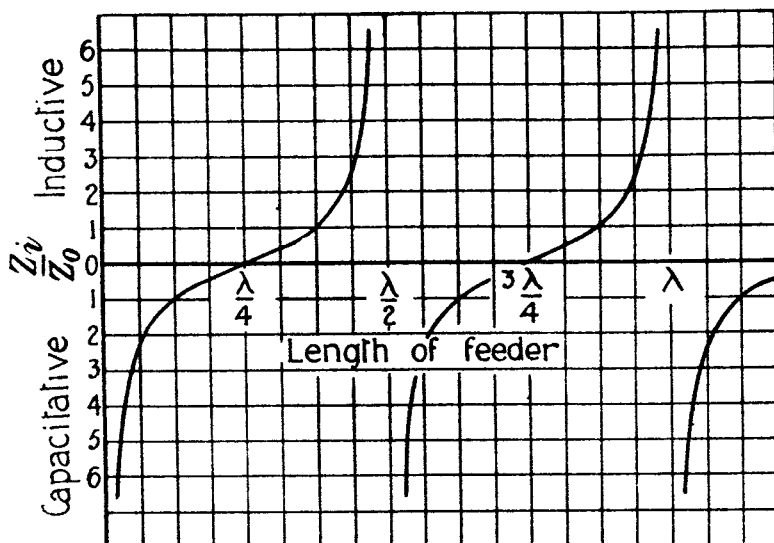


FIG. 36, CHAP. XV.—Reactance of open-circuited line.

(iii) Lengths of line are often used in this manner in matching an aerial array to a transmission line. It is obviously desirable, as a rule, to use the shortest possible lengths of wire, so that in practice a length less than $\frac{\lambda}{4}$ is generally employed. It must be short-circuited if required to act inductively, and on open circuit if required to act capacitively.

Properties of quarter-wave line

87. The properties of a line exactly $\frac{\lambda}{4}$ in length are of particular importance. The input impedance of a loss-free $\frac{\lambda}{4}$ line, terminated by a non-reactive impedance Z_r , is given by

$$\begin{aligned} Z_i &= Z_0 \frac{Z_r \cos \frac{\pi}{2} + j Z_0 \sin \frac{\pi}{2}}{Z_0 \cos \frac{\pi}{2} + j Z_r \sin \frac{\pi}{2}} \\ &= Z_0 \frac{j Z_0}{j Z_r}, \text{ because } \cos \frac{\pi}{2} = 0, \sin \frac{\pi}{2} = 1. \\ \therefore Z_i &= \frac{Z_0^2}{Z_r}. \quad \dots \dots \dots (32) \end{aligned}$$

This property of the $\frac{\lambda}{4}$ line is used for matching purposes. Suppose we have a 100 ohm load, fed from a 600 ohm line. Then $Z_r = 100$, $Z_0 = 600$, and if they are directly connected, reflection will

occur at the termination. To avoid this, we may interpose a $\frac{\lambda}{4}$ length of feeder of such a spacing that its surge impedance Z_m is equal to $\sqrt{Z_0 Z_r}$ i.e. to $\sqrt{100 \times 600} = 245$ ohms. The 600 ohm line will then be correctly terminated, for the input impedance of the $\frac{\lambda}{4}$ line, terminated by 100 ohms, is $\frac{Z_m^2}{Z_r} = \frac{245^2}{100} = 600$ ohms.

Example

In the instance cited above, calculate the spacing of the $\frac{\lambda}{4}$ line in order that $Z_m = 245$ ohms, if the wire is 18 s.w.g.

Diameter of 18 s.w.g. wire is .048 inch, i.e. $r = .024$.

$$Z_0 = 276 \log_{10} \frac{D}{r}$$

$$\log_{10} \frac{D}{r} = \frac{245}{276} = 0.888$$

$$\text{Antilog } 0.888 = 7.727$$

$$\therefore \frac{D}{r} = 7.727$$

$$D = 7.727 \times .024$$

$$= 0.185 \text{ inch.}$$

It is not practicable to space wires as closely as this, except possibly in the case of feeders connected to receiving aerials. A possible solution is a multiple-wire transformation feeder.

Properties of lines of length $\frac{n\lambda}{2}$.

88. We will now consider the input impedance of a length of line equal to some integral multiple of $\frac{\lambda}{2}$, terminated by a non-reactive impedance Z_r . Then

$$Z_i = Z_0 \frac{Z_r \cos \frac{2\pi}{\lambda} l + j Z_0 \sin \frac{2\pi}{\lambda} l}{Z_0 \cos \frac{2\pi}{\lambda} l + j Z_r \sin \frac{2\pi}{\lambda} l} \quad \dots \dots \dots (33)$$

Putting $l = n \frac{\lambda}{2}$, we see that Z_i depends upon whether n is even or odd. If n is odd, $\frac{2\pi}{\lambda} l = n\pi = 3\pi, 5\pi$, etc., and $\sin n\pi = 0$, $\cos n\pi = -1$. Hence $Z_i = Z_0 \times \frac{Z_r}{Z_0} = Z_r$. The magnitude of the voltage at the output terminals is

$$\begin{aligned} V_r &= V_i \frac{Z_r}{j Z_0 \sin \frac{2\pi}{\lambda} l + Z_r \cos \frac{2\pi}{\lambda} l} \\ &= V_i \frac{Z_r}{-Z_r} \\ &= -V_i. \end{aligned}$$

CHAPTER XV.—PARAS. 89–90

Thus we have the important result that a length of line equal to an odd multiple of $\frac{\lambda}{2}$ is a perfect unity-ratio transformer, the output P.D. being equal in magnitude to the input voltage, with a phase difference of 180° . On the other hand if n is even we have $\sin n\pi = 0$, $\cos n\pi = +1$. Hence

$$\begin{aligned} V_r &= V_i \frac{Z_r}{j Z_0 \sin n\pi + Z_r \cos n\pi} \\ &= V_i \frac{Z_r}{Z_r} \\ &= V_i. \\ \text{Also } Z_l &= Z_0 \frac{Z_r + 0}{Z_0 + 0} \\ &= Z_r \end{aligned}$$

Thus, a length of line equal to an even multiple of $\frac{\lambda}{2}$ is a perfect unity-ratio transformer, the input and output P.D.'s being in phase.

Voltage distribution along a feeder

89. The voltage distribution along a feeder may be calculated from the formulae given in previous paragraphs. Taking a length of feeder terminated by a non-reactive impedance $Z_r = Z_0$, we may apply equation (13) of paragraph 74.

$$V_x = V_i e^{-Px} = V_i e^{-j \frac{2\pi}{\lambda} x}.$$

Thus if

$$\begin{aligned} V_i &= \mathcal{V} \cos(\omega t + \varphi) \\ V_x &= \mathcal{V} \cos\left(\omega t + \varphi - \frac{2\pi}{\lambda} x\right), \end{aligned}$$

i.e. the amplitude of the voltage is the same all along the feeder because we have assumed the attenuation to be zero. The phase changes continuously, so that points $\frac{\lambda}{2}$ apart are in opposite phase. The input current is $\frac{\mathcal{V}}{Z_0} \cos(\omega t + \varphi)$ but at a distance x from the input end the current is $\frac{\mathcal{V}}{Z_0} \cos\left(\omega t + \varphi - \frac{2\pi}{\lambda} x\right)$. Since ammeters and voltmeters do not measure phase difference, such a meter will indicate the same R.M.S. current (or voltage) at all points along the line.

90. We will now consider a general case, in which the feeder is terminated by an impedance $m Z_0$ where m may have any positive finite value, either integral or fractional. Applying equations (23) to (27) of paragraphs 80 *et seq.*

$$\begin{aligned} V_x &= V_i \left\{ \frac{m Z_0 \cos \frac{2\pi}{\lambda} (l - x) + j Z_0 \sin \frac{2\pi}{\lambda} (l - x)}{m Z_0 \cos \frac{2\pi}{\lambda} l + j Z_0 \sin \frac{2\pi}{\lambda} l} \right\} \\ &= V_i \left\{ \frac{m \cos \frac{2\pi}{\lambda} (l - x) + j \sin \frac{2\pi}{\lambda} (l - x)}{m \cos \frac{2\pi}{\lambda} l + j \sin \frac{2\pi}{\lambda} l} \right\} \dots \dots \dots (34) \end{aligned}$$

We wish to find how V_x varies with x , and therefore need consider only the numerator of the bracketed portion. This is complex, and its modulus is $\sqrt{m^2 \cos^2 \theta + \sin^2 \theta} = N(V_x)$, and $\theta = \frac{2\pi}{\lambda}(l - x)$.

91. The nature of the variation of V_x at different points in the line, can therefore be obtained by plotting $N(V_x)$ against x . Its maxima and minima may also be obtained by the differential calculus. Differentiating $N(V_x)$ with respect to θ , and equating to zero, we find that maxima or minima are given by $(1 - m^2) \sin \theta \cos \theta = 0$.

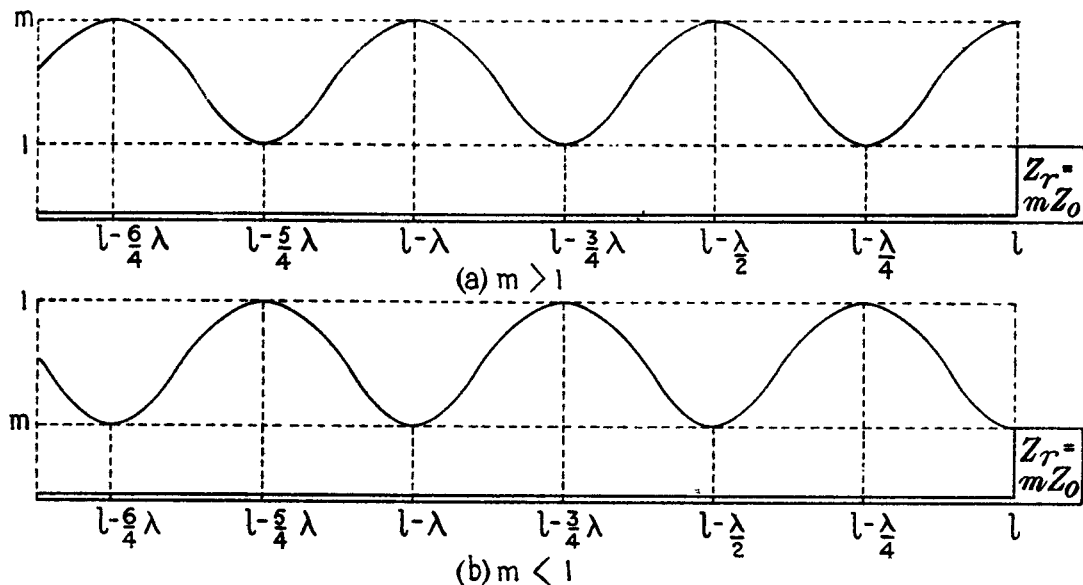


FIG. 37, CHAP. XV.—Location of voltage maxima for values of m greater and less than unity.

Unless $m = 1$, $\sin \theta \cos \theta$ must therefore equal zero, which is the case if θ is any multiple of $\frac{\pi}{2}$ -radians. Hence the maxima or minima occur when $\theta = \frac{n}{2}\pi$ where n is any positive integer or zero. That is, when $\frac{2\pi}{\lambda}(l - x) = \frac{n}{2}\pi$, or $(l - x) = \frac{n\lambda}{4}$, or zero. It follows, therefore, that either a maximum or a minimum of voltage will occur at the termination, i.e. where $l = x$. If m is greater than unity it will be a maximum, if m is less than unity, a minimum. At a distance of $\frac{\lambda}{4}$ from the termination, there will be another turning point so that the voltage distribution will be either as in fig. 37a or fig. 37b, depending on whether $m > 1$ or $m < 1$. It will be seen that m is the ratio of the maximum to the minimum P.D. or *vice versa*. The current distribution along the feeder may be calculated in a similar manner. The resulting curves are very nearly sinusoidal but not exactly so, except in the case of short-circuited or free lines, because a terminal load will necessarily call for a feed current. The calculated current distribution for a 600 ohm line, for ratios $\frac{I_{\min}}{I_{\max}}$ from 0.1 to 0.9, are given in fig. 38.

CHAPTER XV.—PARA. 92

Measurement of surge impedance

92. (i) If the operating frequency is sufficiently low, it is possible to determine the surge impedance of a line by actual measurement. Let the length of the line be l . The input impedance is first measured, with the receiving end on open circuit; let this be Z_f (f for "free"). The input impedance Z_c (c for "closed") with the receiving end on short-circuit is also found. Then

$$Z_f = Z_0 \coth Pl$$

$$Z_c = Z_0 \tanh Pl$$

from paragraphs 78 and 79.

It follows that

$$Z_f Z_c = Z_0 \coth Pl \times Z_0 \tanh Pl$$

$$= Z_0^2$$

$$\therefore Z_0 = \sqrt{Z_f Z_c}.$$

(ii) The following method has also been proposed. On erection, the line is extended for rather more than $\frac{\lambda}{8}$ past the proposed terminating point, and is then energized at the intended frequency, or a closely adjacent one, in such a manner that stationary waves are set up along the line. The wavelength on the line itself is obtained by observation of the minimum current at adjacent current nodes by means of an ammeter and transformer (see paragraphs 4 and 128). A $\frac{\lambda}{8}$ length of line is then removed from the free end and a calibrated variable condenser joined across the ends of the line in its place. The capacitance is varied as necessary until the current minima appear at the same points as before; when this is achieved, the capacitance is exactly equivalent to the length of line which was removed. Referring to paragraph 79, the impedance of a $\frac{\lambda}{8}$ length of open loss-free line is

$$Z_{\frac{\lambda}{8}} = -jZ_0 \cot \frac{2\pi}{\lambda} l$$

$$= -jZ_0 \cot \frac{2\pi}{\lambda} \frac{\lambda}{8}$$

$$= -jZ_0 \cot \frac{\pi}{4}$$

$$= -jZ_0.$$

Since the capacitance C has exactly the same effect on the line as the impedance $Z_{\frac{\lambda}{8}}$ it follows that

$$\frac{1}{j\omega C} = Z_{\frac{\lambda}{8}}$$

$$\frac{1}{j\omega C} = -jZ_0$$

or

$$Z_0 = \frac{1}{\omega C}.$$

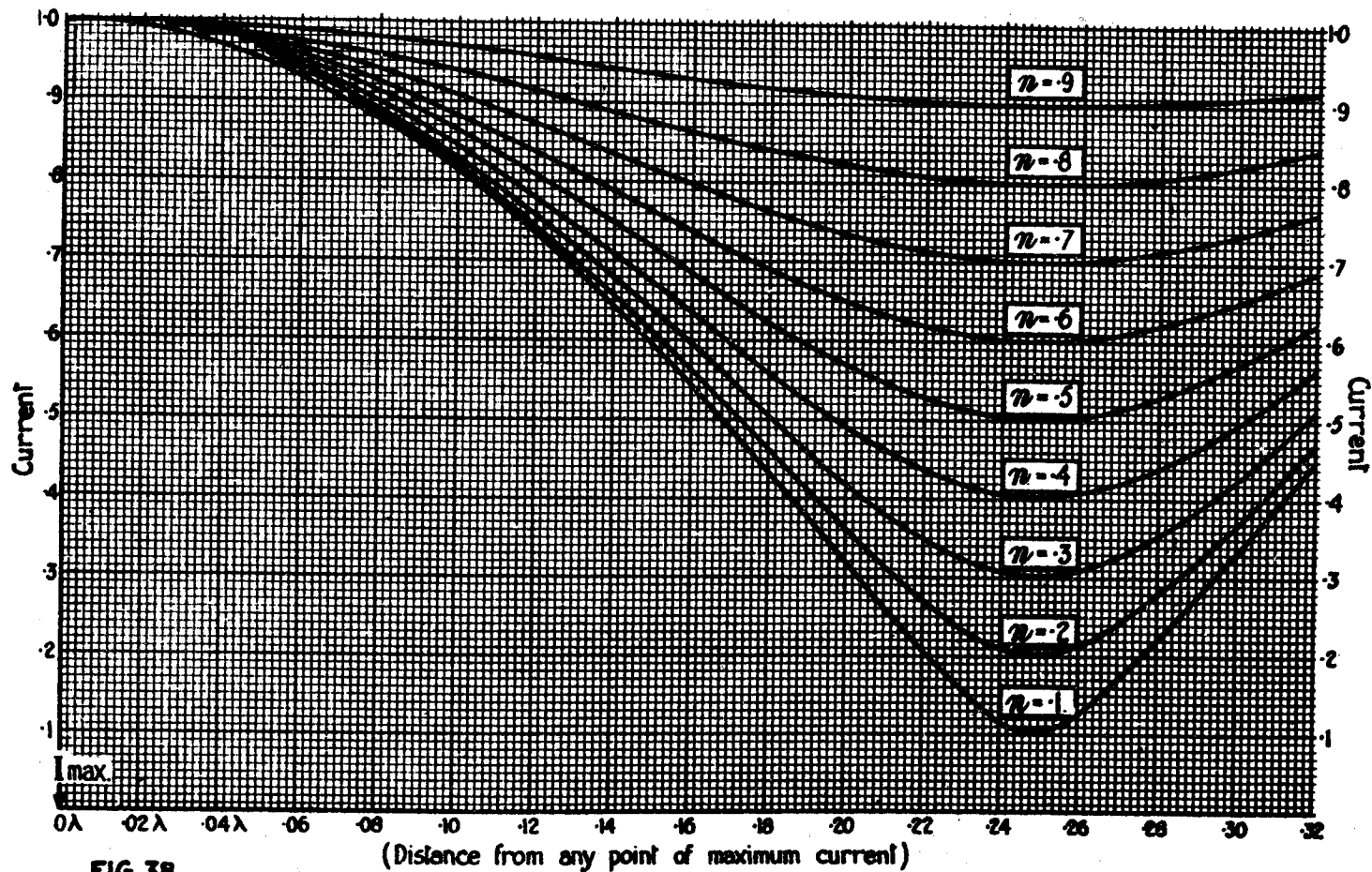


FIG.38
CHAP.XV

CURRENT DISTRIBUTION ALONG 600 OHM FEEDER

Radiation due to travelling wave

93. Although in most forms of aerial the arrangement is such that stationary waves are established along the wires, it must not be thought that this is an essential requirement for radiation to occur. The fact that radiation can and does occur from wires carrying travelling waves is of importance from two points of view. First, in the case of properly terminated feeder lines, considerable radiation will occur unless the lines are very close together, i.e. less than about 0.05λ. Second, it is possible to design aerial arrays for directional transmission and reception, the action depending entirely upon the radiating properties of a long, properly terminated wire. The directivity of such an aerial for receiving purposes depends upon the reciprocal properties mentioned in paragraph 3.

94. If a long straight wire is situated in free space and carries a travelling current wave, the radiation field set up by the current is easily calculated. Referring to fig. 39 consider a wire of length l and let the current at an origin O at the mid-point of the length be $I_0 = I e^{j\omega t}$. At a

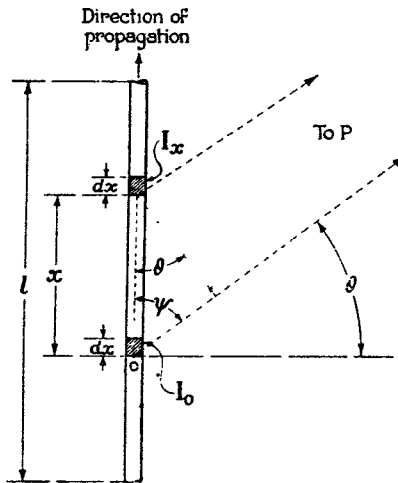


FIG. 39, CHAP. XV.—Wire carrying travelling wave.

point P , at a distance r ($\gg l$) from the origin, and at an angle θ to the perpendicular through O , the field $d\gamma_0$ due to the current in a short length dx of conductor closely adjacent to O , may be found by treating the length dx as a hertzian doublet, giving

$$d\gamma_0 = \frac{60}{r} \frac{\pi}{\lambda} \cos \theta \cdot dx \times I e^{j\left(\omega t + \frac{\pi}{2} - \frac{2\pi}{\lambda} r\right)} \quad \dots \quad (35)$$

Consider another element of length dx at a distance x from the origin. Since the wire carries a travelling wave, the current I_x in this element will have the same amplitude as at the origin, but will be out of phase with it. If x is measured in the direction of propagation along the wire, I_x

lags on I_0 by $\frac{2\pi}{\lambda} x$ radians. Also, the point P is distant $(r - x \sin \theta)$ from the element, and the field due to the latter will be

$$\begin{aligned} d\gamma &= \frac{60}{r} \frac{\pi}{\lambda} \cos \theta \cdot dx \times I_x e^{j\left[\omega t + \frac{\pi}{2} - \frac{2\pi}{\lambda} (r - x \sin \theta)\right]} \\ &= \frac{60}{r} \frac{\pi}{\lambda} \cos \theta \cdot dx \times I e^{j\left(\omega t + \frac{\pi}{2} - \frac{2\pi}{\lambda} r\right)} e^{j \frac{2\pi}{\lambda} x \sin \theta} \quad \dots \quad (36) \end{aligned}$$

CHAPTER XV.—PARA. 95

But

$$\begin{aligned} I_x &= I_0 \varepsilon^{-j \frac{2\pi}{\lambda} x} \\ \therefore d\gamma &= \frac{60}{r} \frac{\pi}{\lambda} \cos \theta I_0 \varepsilon^{j \left(\frac{\pi}{2} - \frac{2\pi}{\lambda} r \right)} \varepsilon^{j \frac{2\pi}{\lambda} x (\sin \theta - 1)} dx, \\ &= A \cos \theta \varepsilon^{j \frac{2\pi}{\lambda} x (\sin \theta - 1)} dx \end{aligned} \quad \dots \quad (37)$$

where, for brevity, $A = \frac{60}{r} \frac{\pi}{\lambda} I_0 \varepsilon^{j \left(\omega t + \frac{\pi}{2} - \frac{2\pi}{\lambda} r \right)}$.

The total field set up by the whole length of conductor is obtained by integrating between the limits $x = +\frac{l}{2}$ and $x = -\frac{l}{2}$, giving

$$\begin{aligned} \gamma &= A \cos \theta \int_{x = -\frac{l}{2}}^{x = +\frac{l}{2}} \varepsilon^{j \frac{2\pi}{\lambda} x (\sin \theta - 1)} dx \\ &= \frac{A \cos \theta}{j \frac{2\pi}{\lambda} (\sin \theta - 1)} \left[\varepsilon^{j \frac{\pi l}{\lambda} (\sin \theta - 1)} - \varepsilon^{-j \frac{\pi l}{\lambda} (\sin \theta - 1)} \right] \\ &= -j \frac{\lambda A \cos \theta}{2\pi (\sin \theta - 1)} \times 2 \sin \left\{ \frac{\pi l}{\lambda} (\sin \theta - 1) \right\} \quad \dots \quad (38) \end{aligned}$$

Polar diagrams

95. Neglecting the factors $-j$ and A for the present, the field varies with the angle θ in accordance with the equation

$$f(\theta) = \frac{\cos \theta}{\sin \theta - 1} \sin \left[\frac{\pi l}{\lambda} (\sin \theta - 1) \right],$$

and this is more compactly expressed in terms of the angle $\psi = \frac{\pi}{2} - \theta$ which is the angle of the direction r with reference to the axis of the wire. Then

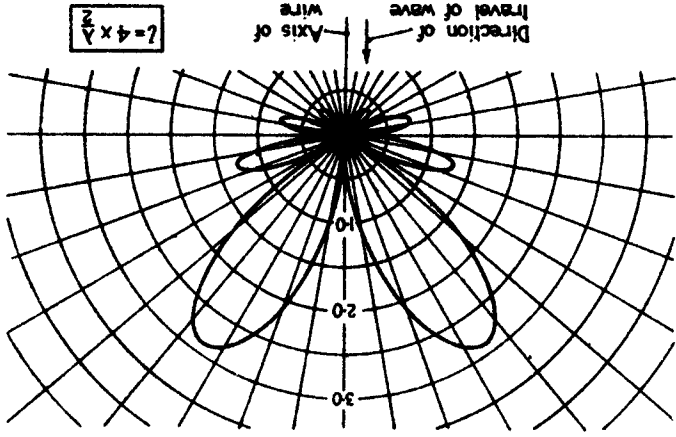
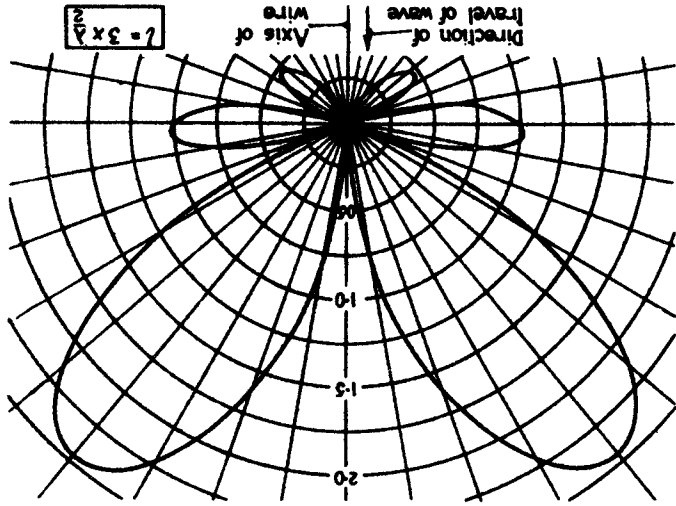
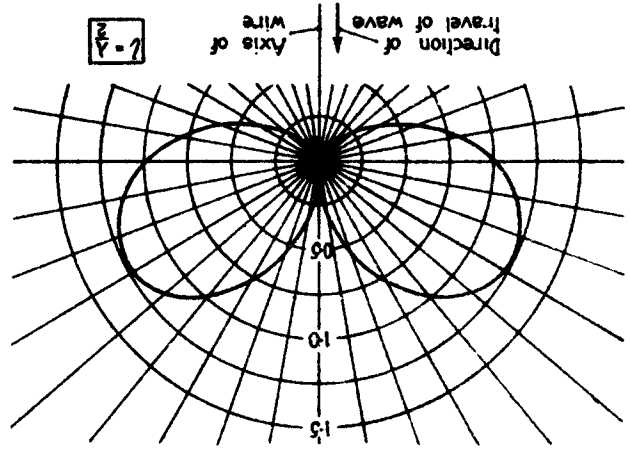
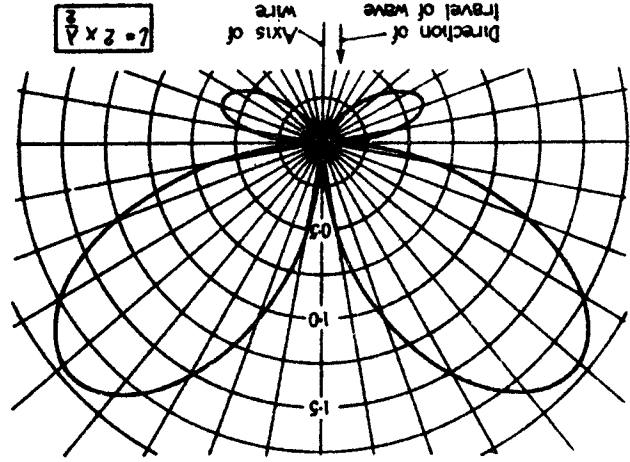
$$f(\psi) = \cot \frac{\psi}{2} \sin \left[\frac{2\pi l}{\lambda} \sin^2 \frac{\psi}{2} \right].$$

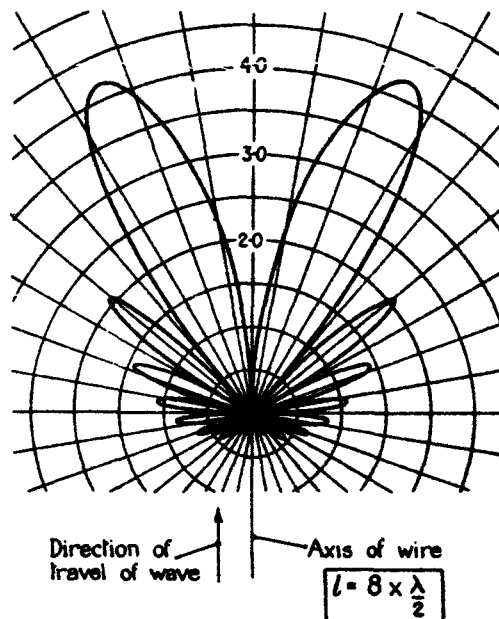
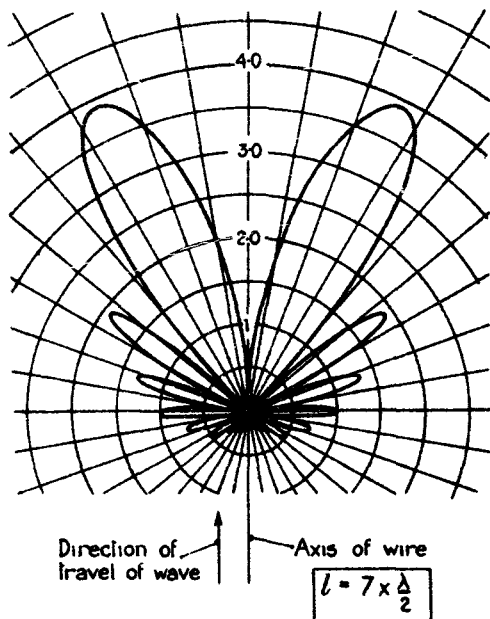
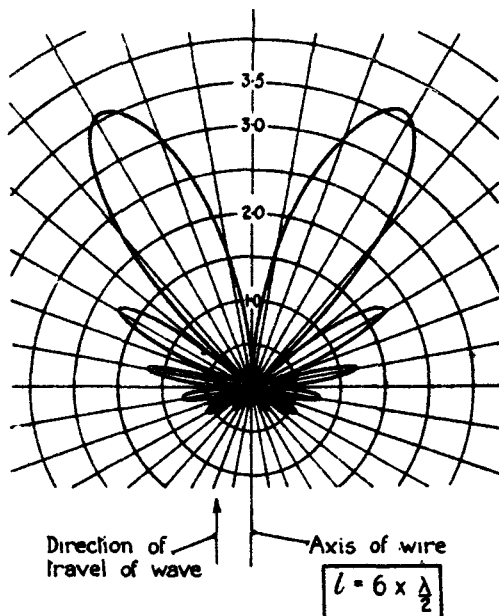
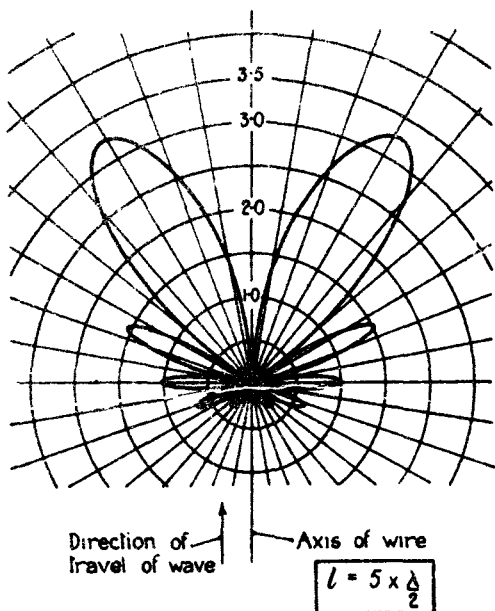
The total instantaneous field is therefore

$$\begin{aligned} \gamma &= \frac{60}{r} \frac{\pi}{\lambda} I_0 \varepsilon^{j \left(\omega t + \frac{\pi}{2} - \frac{2\pi}{\lambda} r \right)} \left\{ -j \frac{\lambda}{\pi} \cot \frac{\psi}{2} \sin \left[\frac{2\pi l}{\lambda} \sin^2 \frac{\psi}{2} \right] \right\} \\ &= \frac{60}{r} I_0 \varepsilon^{j \left(\omega t - \frac{2\pi}{\lambda} r \right)} \left\{ \cot \frac{\psi}{2} \sin \left[\frac{2\pi l}{\lambda} \sin^2 \frac{\psi}{2} \right] \right\} \quad \dots \quad (39) \end{aligned}$$

It is interesting to note that this field lags by $\frac{\pi}{2}$ radians on that which would be produced by a stationary wave in the same wire. The portion enclosed in curved brackets, that is, $f(\psi)$, is plotted

FIELD STRENGTH DUE TO STRAIGHT WIRE CARRYING TRAVELLING WAVE





FIELD STRENGTH DUE TO STRAIGHT WIRE
CARRYING TRAVELLING WAVE

in fig. 40, sheets 1 and 2, for various values of l up to 4λ . It is seen that the length of wire has an influence upon the magnitude of the maximum radius vector and also upon the angle of the main lobe with reference to the axis of the wire. These are collected and shown graphically in fig. 41.

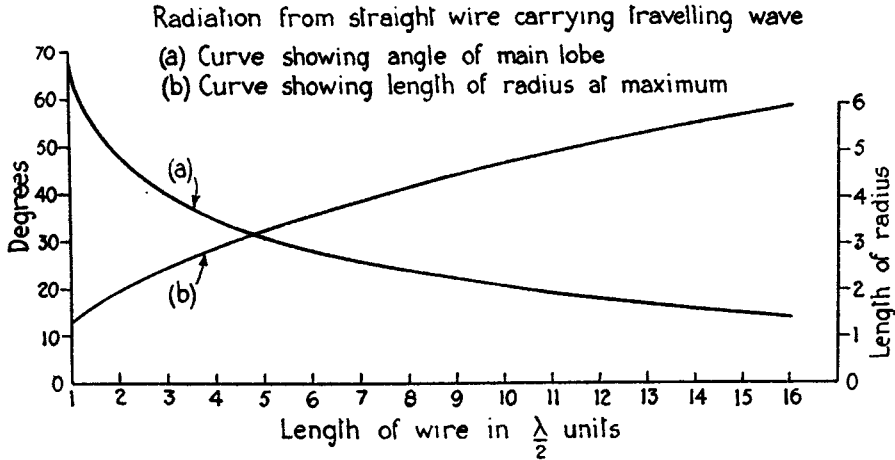


FIG. 41, CHAP. XV.—Radiation due to travelling wave.

96. It is of interest to refer to the particular case when $l = \frac{\lambda}{2}$. The R.M.S. field is then

$$r_{(l = \lambda/2)} = \frac{60}{r} I \cot \frac{\psi}{2} \sin \left(\pi \sin^2 \frac{\psi}{2} \right)$$

As previously shown, a dipole gives a field

$$r(\text{dipole}) = \frac{60}{r} I \frac{\cos \left(\frac{\pi}{2} \sin^2 \varphi \right)}{\cos \varphi}$$

which is equal to $\frac{60}{r} I$ when $\varphi = 0$. For a length $l = \frac{\lambda}{2}$ the maximum value of $f(\psi)$ occurs when

$\psi \doteq 65^\circ$. This value of $f(\psi)$ is nearly 1.25, so that a $\frac{\lambda}{2}$ length of conductor carrying a travelling wave sets up a field 25 per cent. greater than a dipole carrying the same current. In the former case the current is of course the same at all points in the line, whereas in the dipole the current referred to is that at the mid-point.

Application to radiation from transmission line

97. Care must be taken when applying the above results to transmission lines. Let us suppose a twin wire transmission line to have such small separation that the two wires may be regarded as coincident in space. Each wire has a current I and carries a travelling wave; the polar diagram will thus be the algebraic sum of those corresponding to the respective waves. The currents in the two wires are in opposite directions but the wave direction is the same. Hence the two diagrams lie upon each other but are of opposite sign, and the feeder is shown to be non-radiative. If, however, the wires are energized in such a manner that both the wave direction and the

CHAPTER XV.—PARAS. 98-100

instantaneous current are reversed in direction, the polar diagram is the sum of two diagrams appropriate to the length of conductor, one of which is turned upside down and so placed that the two origins coincide. This is easily seen by tracing the diagram for $l = \frac{\lambda}{2}$ (fig. 40) on tracing paper, turning it about the origin through 90° , and then adding the polar radii of the tracing and the original diagram. The result is found to be identical in shape with fig. 6, but has a maximum radius of two units. This is not surprising since fig. 6 is by hypothesis the polar diagram of a conductor carrying one half of a stationary wave formed by reflection at its open ends, the loop current being $2I$. By suitable manipulation, then, the travelling wave diagrams may be used to determine the polar diagrams of conductors carrying stationary waves.

98. Reverting to the case of a single wire carrying a travelling wave it is obvious that, in free space, the polar diagram in the plane perpendicular to the axis of the wire is a circle, i.e. the wire radiates uniformly in all directions. The solid polar diagram is therefore obtained by rotating the axial diagram about the wire. It follows that if we have two parallel wires carrying travelling waves, the polar diagram of the two is obtained by multiplying the axial diagram by the appropriate Grating Factor. In a twin wire transmission line carrying equal and opposite currents, the grating factor is given by row E of fig. 8. These diagrams are however of little practical use, for the present purpose, because transmission line spacing is usually much less than $\frac{\lambda}{8}$, the smallest spacing given. For closer spacing, however, the grating factor diagram closely approximates to two circles in contact at the origin; as the Grating Factor is $2 \cos \left(90 - \frac{180d}{\lambda} \cos \theta \right)$, the diameter of this circle is easily seen to be $2 \sin \frac{180d}{\lambda}$. For example, if $\frac{d}{\lambda} = 0.05$, $\frac{180d}{\lambda} = 9^\circ$, $2 \sin 9^\circ = 0.3128$ which is the diameter of each circle. It will be seen that for this spacing there is quite appreciable radiation, but this is greatly reduced as the spacing is decreased.

Effect of unbalanced currents in transmission line

99. When the spacing is very small, i.e. of the order of $\frac{\lambda}{50}$, the effect of unbalanced currents is of more importance than the actual spacing. By methods already used it is easily shown that if the currents are I_A , I_B , and $I_B = M/\beta I_A$, the field at a radius r and angle ψ is approximately

$$r = \frac{60}{r} f(\psi) I_A \sqrt{1 + M^2 + 2M \cos \beta}.$$

Thus, if $\frac{d}{\lambda}$ is very small, and $\beta = 180^\circ$, the Grating Factor diagram becomes a circle of diameter $1 - M$.

Effect of proximity of ground

100. So far the presence of the ground below the feeder line has been neglected. If perfect conductivity is assumed, the image of the feeder must be considered to carry a wave travelling in the same direction as in the feeder, but with the instantaneous current in the opposite direction at all points. The field due to the feeder must be therefore multiplied by the Vertical Distribution Factor appropriate to a horizontal dipole at a distance above ground equal to that of the feeder.

Properties of twin wire and concentric feeders

101. In order that the transmission line theory may hold, it is necessary that the current at each point in one of a pair of twin wires shall be equal in magnitude to the current at the corresponding point in the other. Now each wire has a capacitance with respect to earth, and there is also a capacitance between the two wires. The line currents at corresponding points can only be of equal magnitude if all corresponding points have equal and opposite voltages with respect to earth, and the currents, although equal in magnitude, are then exactly 180° out of phase. Under these conditions the line is said to be balanced with respect to earth. In a properly terminated and balanced line, the power losses are almost entirely due to the ohmic resistance, and the efficiency of transmission fairly high. As an approximation, the efficiency

may be taken as $\left(100 - \frac{2l}{\lambda}\right)$ per cent., l being the length of the line. A little reflection will show

that although the line itself may be balanced, the circuit as a whole cannot be so, unless the input and output impedances are also symmetrical with respect to earth. Thus, a horizontal dipole (fig. 42a) is a suitable load for a twin wire feeder, but an earthed aerial (fig. 42b) is not. If it is necessary to feed the latter by means of a twin wire feeder, a coupled circuit may be used, as in fig. 42c. It may be necessary to place an electro-static screen between the two coils as shown. As an alternative the circuit of fig. 42d is suggested. Here the electrical centre of the

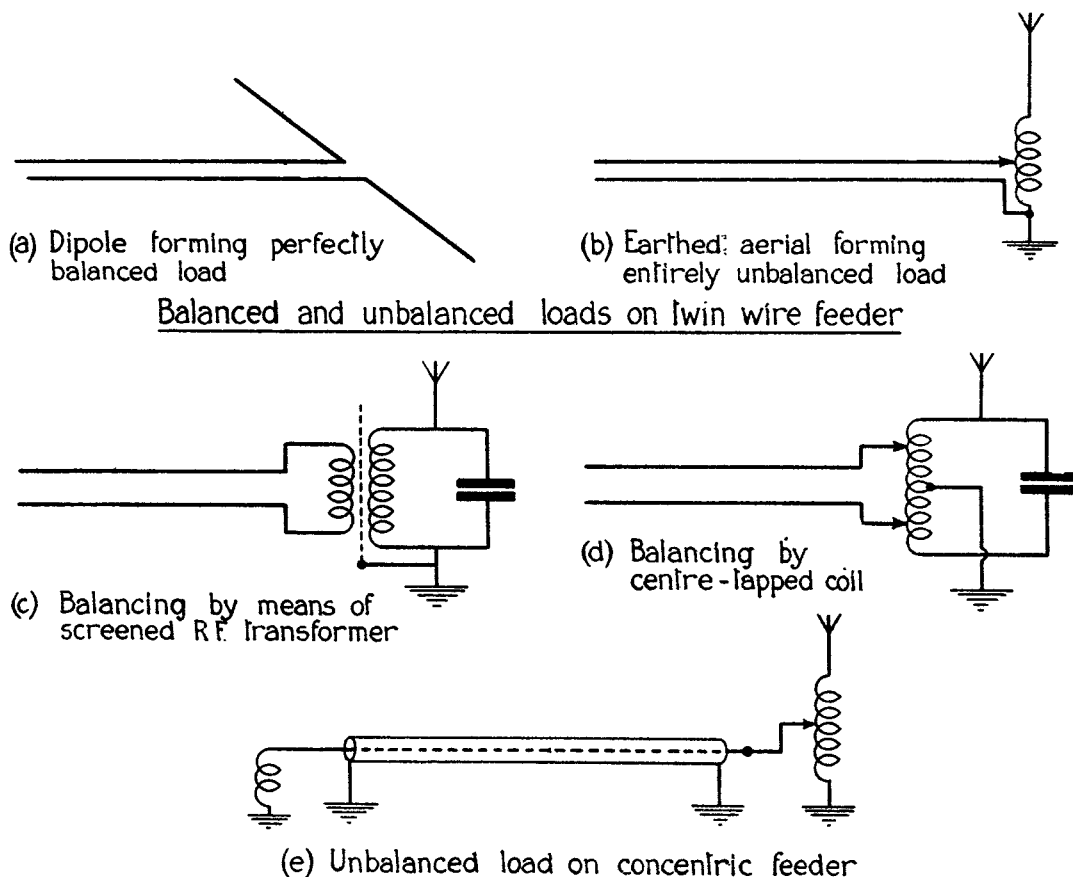


FIG. 42, CHAP. XV.—Balanced and unbalanced loads.

CHAPTER XV.—PARAS. 102-103

coil is earthed, and the feeders are tapped in at electrically equidistant points on each side of earth. It must be noted that since the aerial is connected to one end of the coil, the capacitance to earth of its two ends may be very different, and the electrical and geometrical centres are not usually coincident. Similar considerations apply to the input end of the feeder. Although it is possible to transmit power along an unbalanced line (for example, as in fig. 42b) it is found that standing waves are set up, with maxima and minima of different values and at non-corresponding points in the two wires. The efficiency of transmission is then very low and, in addition, it is impossible to predict, even approximately, the behaviour of an unbalanced line.

102. In a concentric feeder there is practically no external field, because the currents in the outer conductor are confined to a very thin layer on the inner surface; and the outer portions act merely as a screen. The outer conductor may, therefore, be earthed without affecting the electrical characteristics as a transmission line. It follows, therefore, that an earthed aerial may be fed by means of a concentric feeder as in fig. 42e. On the other hand, if it is required to feed a balanced load from a concentric feeder, some form of coupling device must be employed. This is exactly opposite to the conditions governing the use of twin wire feeders.

Methods of balancing the concentric feeder

103. As it is often necessary to feed a balanced aerial or aerial array by means of a concentric feeder, two methods of doing so will be described. The first is very simple and depends upon the fact that a length $\frac{\lambda}{2}$ of transmission line acts as a perfect 1/1 transformer (with a phase reversal of 180°). Referring to fig. 43a, suppose T_1, T_2 to be the input terminals and T_3, T_4 the output terminals of a section of feeder $\frac{\lambda}{2}$ in length. If an impedance Z_r is connected across T_3, T_4 , and a P.D. V_i exists at T_1, T_2 , the voltage across Z_r at T_3, T_4 , will be $-V_i$. In addition, the impedance of the line, as measured at the terminals T_1, T_2 , with the load Z_r so connected, will be Z_r ohms. The load so connected is, however, completely unbalanced. Now suppose that at the end of a transmission line, we measure backwards towards the input end a distance equal to $\frac{\lambda}{2}$, and bring out a suitably insulated connection from the inner conductor, as shown in fig. 43b. The terminals T_3 and T_4 are then at equal and opposite potentials with respect to earth, and are suitable for

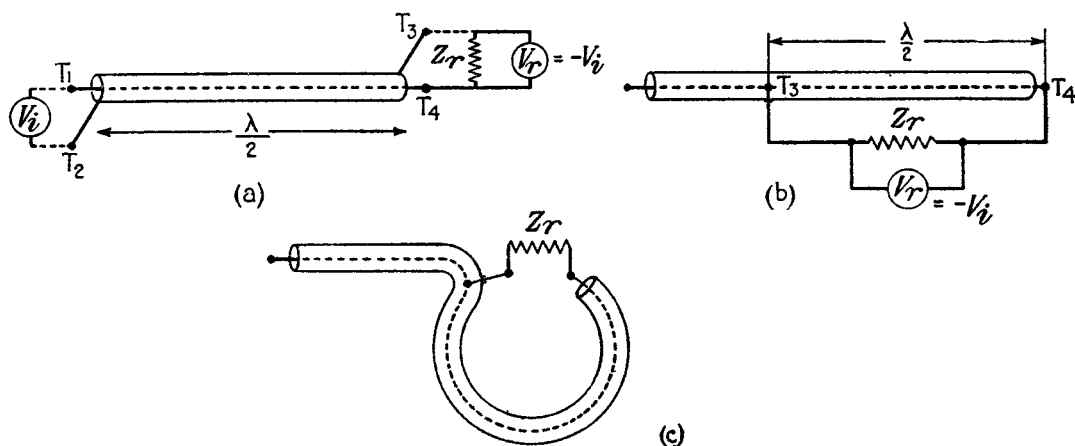


FIG. 43, CHAP. XV.—Balanced output from concentric feeder—first method.

feeding a balanced load. Since, however, T_3 and T_4 are usually required to be in proximity it is convenient to fold the $\frac{\lambda}{2}$ length of cable as in fig. 43c. This will probably affect the velocity of the wave along this portion and it may be necessary to determine the exact length by trial and error.

104. The effective impedance of the load, when connected to the output terminals, is only $\frac{Z_r}{4}$.

This may be seen from the following considerations. Since the actual impedance between T_3 and T_4 is Z_r , and it is balanced with respect to earth, its centre point is at or near earth potential. The outer conductor is also at or near earth potential, so that between T_3 and earth we have an impedance of $\frac{Z_r}{2}$ ohms and an impedance of $\frac{Z_r}{2}$ ohms between T_4 and earth. But the latter impedance may equally be considered to be connected between T_3 and earth, from the argument in the preceding paragraph, and therefore the transmission line must be considered as being terminated at T_3 , by two impedances in parallel, each of $\frac{Z_r}{2}$ ohms, i.e. by $\frac{Z_r}{4}$ ohms.

105. The second method is as follows. A length l of copper tube A B of the same external diameter as the outer conductor of the feeder, is placed parallel to the end of the transmission line, and is electrically connected to the latter at A as in fig. 44a. The inner conductor of the

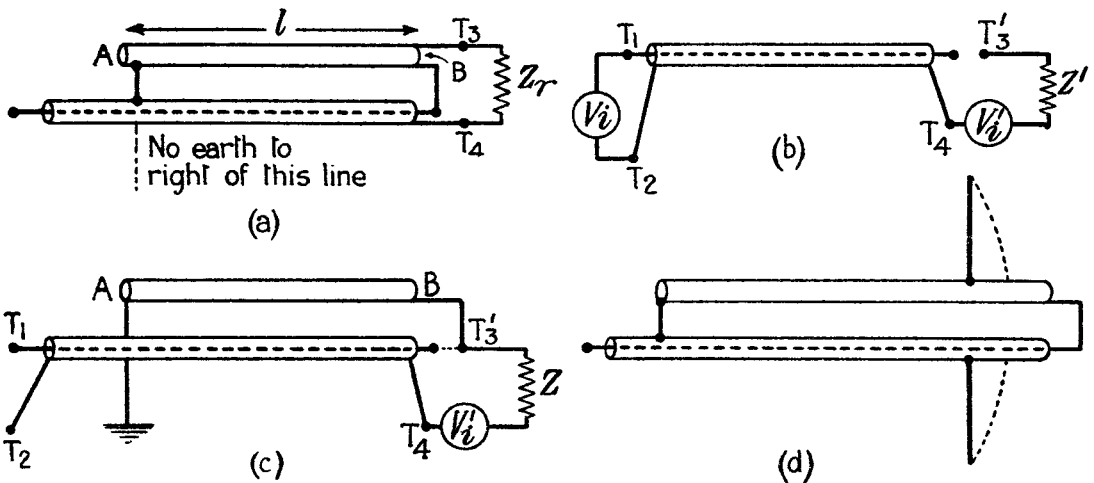


FIG. 44, CHAP. XV.—Balanced output from concentric feeder—second method.

transmission line is connected to the added tube at B and the balanced output is then taken from the terminals T_3 T_4 , which are connected to the external conductor of the line and the added conductor respectively. It must be noted that the length l of both these conductors must be insulated from and preferably symmetrically disposed with respect to earth. The explanation of the operation of this device is simple, but rather more difficult to visualize than that previously described. Suppose that at the input end we apply a voltage V_i , it is possible to find an equivalent voltage V_i' and impedance Z' which, when connected between the terminals T_3' T_4 , will cause the same P.D. at these terminals and will deliver the same current to the load. Now the point T_3' has (practically) zero capacitance to earth because it is entirely shielded by the outer conductor. It may therefore, as a preliminary, be considered as an entirely isolated point connected to T_4 by a generator of voltage V_i' and an impedance Z' in series, neither of which possess capacitance

CHAPTER XV.—PARA. 106

with respect to earth (fig. 44b). If now the extra length of conductor is added as in fig. 44c, and the terminal T'_3 connected to the point B, it is obvious that the output impedance is symmetrical with respect to earth. The inner conductor may therefore be connected to T'_3 as in the dotted line. We have, however, added, at the points $T'_3 T_4$ an additional parallel impedance due to the short transmission line formed by the two parallel tubes. Denoting this by Z_2 , it has already been shown that $Z_2 = j Z_0' \tan \frac{2\pi}{\lambda} l$, where Z_0' is the surge impedance of the length l of parallel tube.

If $l = \frac{\lambda}{4}$, Z_2 becomes infinite, while if $l = \frac{\lambda}{2}$, Z_2 is zero. On either side of $\frac{\lambda}{4}$, Z_2 is either capacitive or inductive, hence by suitable choice of l an effective reactance of any desired value may be placed in parallel with the actual load. The resistance of the load may be matched to the surge impedance of the feeder by choosing suitable locations for the terminals T'_3 , T_4 ; in fig. 44d, a $\frac{\lambda}{2}$ dipole is fed in this way, the output terminals being located a short distance from the ends of the parallel tubes.

Comparison of twin-wire and concentric lines

106. In practice, both twin-wires and concentric feeders are used according to local conditions. It is not possible to give any definite rules which will govern the adoption of either type. The following summary of their relative advantages and disadvantages should be taken into account in any decision.

(i) Type of load

(a) Twin feeders are inherently balanced and are suitable for any type of load consisting of an arrangement of dipoles. If the load is not symmetrical with respect to earth, some form of coupling device must be adopted.

(b) Concentric feeders are inherently unbalanced and are suitable for unbalanced loads such as earthed aerials. If the load is symmetrical with respect to earth some form of coupling device must be adopted.

(ii) Constructional

(a) Twin feeders are cheap to construct and repair. In the field it is even possible to erect a workable line from field telegraph poles and improvised insulators such as glass bottles, although of course a high transmission efficiency cannot be expected. On the other hand, since a twin-wire line should be several feet above the ground, it is difficult to adopt this type where the transmitter or receiver is installed underground.

(b) Concentric feeders are expensive to construct, difficult to repair in the event of a mechanical failure, and are impossible to improvise. They may, however, be buried and led to an underground station.

(iii) Convenience

(a) Twin-wire feeders occupy a considerable space, particularly where a large number of aerials are energized from transmitters in the same building, because they have an external field, and unless different lines are well apart they will affect each other.

(b) Concentric feeders are very compact and may be placed in proximity without mutual interaction.

(iv) Breakdown voltage

(a) The breakdown voltage between twin wires of suitable spacing is very high. Flash over is not likely to occur with properly matched and balanced loads, at any rate with the power required in service transmitters.

(b) With concentric feeders, the spacing between conductors is comparatively small and they are more likely to flash over. It is therefore essential, from this point of view alone, to ensure that no standing waves exist in the feeder.

(v) *Transmission losses*

For the powers used in the service, i.e. up to a few kilowatts, there is probably little to choose between the two types, although accurate figures are not available. For very high power (e.g. 500 kW.) concentric feeders may be better.

(vi) *Radiation and pick-up*

(a) The power radiated by a correctly matched twin-wire feeder is not large. In reception the pick-up is correspondingly small, but is often difficult to eliminate completely.

(b) With concentric feeders, radiation and pick-up are practically non-existent. For special purposes where freedom from pick-up is absolutely essential, e.g. a remote D/F system, concentric feeders must be employed. Again, if it is proposed to use a very high frequency transmitter in an aeroplane, and a feeder line is necessary, the concentric type is almost compulsory.

MATCHING DEVICES

Necessity for matching

107. (i) It will be appreciated from the foregoing that in order to convey the greatest possible amount of power from a transmitter to an aerial system by means of a feeder line, the input impedance of the feeder must be matched to the output impedance of the transmitter, and the input impedance of the aerial must be matched to the surge impedance of the feeder. At the

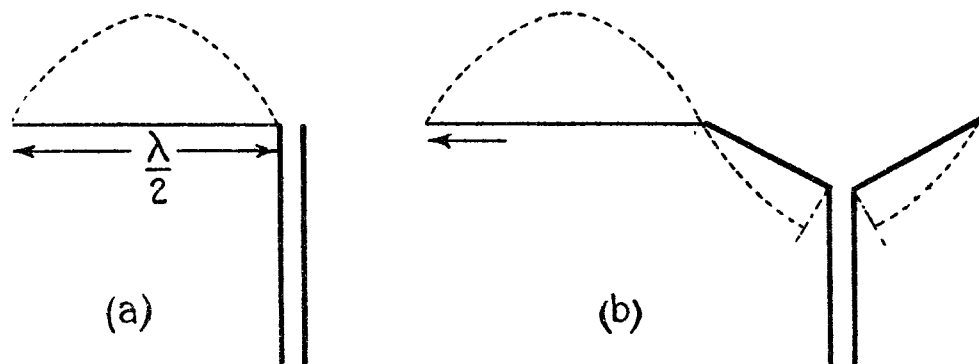


FIG. 45, CHAP. XV.—Example of mis-matching.

transmitter end, the suitable matching devices are usually incorporated in the design of the transmitter and need no further comment. The matching of aerial to feeder must often be dealt with by the personnel responsible for bringing the station into operation, especially on active service. Unless this matching is fairly close, the efficiency of the station may be very low. As an example of what must be avoided if possible, take the arrangement shown in fig. 45a. This is

fairly satisfactory if the feeders are, electrically, $\frac{\lambda}{4}$ in length, as explained in Chapter VII. If, however, a feeder of indefinite length is used, a purely physical consideration will show that the arrangement is far from efficient. If it is considered as a $\frac{\lambda}{2}$ aerial connected to one side of a transmission line, the aerial has an input resistance of some 3,000 ohms, while the feeder will have a surge impedance of the order of 600 ohms. On the other hand, we may assume that the feeder

CHAPTER XV.—PARAS. 108–109

is effectively terminated at a point $\frac{\lambda}{4}$ from the aerial connection. In fig. 45b the final $\frac{\lambda}{4}$ length of feeder has been splayed out in order to show the current distribution on this assumption. It is now seen that the arrangement is equivalent to a λ aerial fed at a current loop, and as the aerial is so folded that a length of $\frac{\lambda}{2}$ is non-radiative, its input resistance is of the order of 80 to 100 ohms only. Regarding the arrangement in either of these ways it is seen that the ratio $\frac{Z_r}{Z_0}$ (or $\frac{Z_0}{Z_r}$) is of the order of 5. Although it is rarely possible to obtain a perfect match ($\frac{Z_r}{Z_0} = 1$), the aim should be to attain this within 25 per cent.

(ii) Since a well-designed radio-frequency feeder is practically loss-free, and its surge impedance is to all intents and purposes purely resistive, stationary waves will only be suppressed if the termination is also purely resistive. If the input impedance of the aerial has a reactive component, this must be balanced out by incorporating an equal and opposite reactance in the termination.

Limitations of R.F. transformer as matching device

108. At first sight the problem of achieving an approximate match between the feeder and the aerial system would appear to be comparatively simple, merely involving the design of a suitable radio-frequency transformer. In practice, however, it is very difficult to obtain a practical solution by this method. This would present little difficulty if it were practicable in a transformer of this kind to achieve a coupling factor approaching unity, but actually it is rarely possible for it to exceed 0.5. This is due to the necessity for well spacing the coils, in order to avoid capacitance coupling and to permit the development of high voltage across the input and output terminals without insulation breakdown. It is highly desirable that, looking into the input terminals of the matching device, the load shall be non-reactive. If this is not so the power factor of the load will be less than unity and, for a given input to the aerial, the line current must be greater than with a non-reactive load. Since the line cannot have zero resistance, this must lead to power loss in the feeder and to low efficiency. Further, the surge impedance of a radio-frequency line is practically non-reactive and should, therefore, be terminated by a purely resistive load. It is possible to bring the power factor to unity by the introduction of a suitable condenser or condensers in addition to the transformer, but the design of such condensers is again beset with difficulty. They must be located near the aerial and, therefore, in weatherproof casing, and yet must be very highly insulated from earth. In some instances, the capacitance may be only about $100\mu F$ and yet the plate area must be sufficient to carry the full feed current without overheating. In addition, consideration of breakdown voltage may necessitate a large spacing although the external field must be negligible. The two latter requirements lead to a very bulky and extremely expensive condenser. Unless a transformer is absolutely essential it is customary to perform the matching by means of an electrical network consisting of arrangements of impedances. These impedances often take the form of suitable lengths of transmission line.

Principle of matching network

109. In dealing with matching by means of electrical networks, we shall assume that the aerial itself at its input terminals, offers resistance only. This resistance will be denoted by R_A . The line will have a surge impedance Z_0 , which may be taken to be purely resistive and denoted by R_0 . The resistances R_A and R_0 being unequal, we require to insert some matching device between R_0 and R_A . This will be some arrangement of reactances which must be of the lowest possible resistance. Before dealing with some of the various possible arrangements consider the matching unit shown in fig. 46, in which the exact arrangement of the apparatus is unknown.

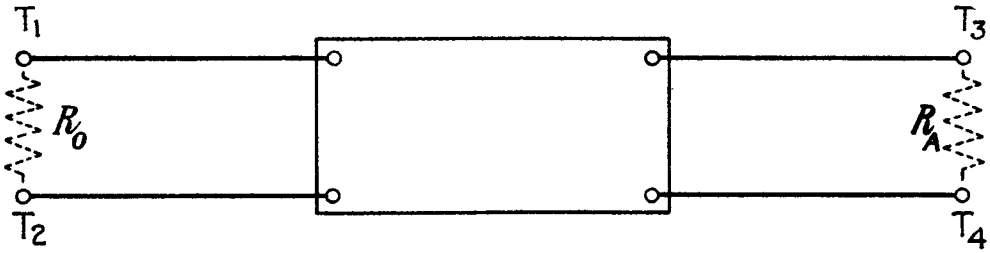


FIG. 46, CHAP. XV.—Insertion of matching network.

The requirements are that if the resistance R_A is connected across the terminals T_3, T_4 , the impedance, measured at the terminals T_1, T_2 , is R_0 . On the other hand, if a resistance R_0 is connected to T_1, T_2 , and the impedance measured at T_3, T_4 , it must be equal to R_A . This reciprocal relation is the essential property of any matching network, and is true if the latter consists of reactances only.

L unit

110. The simplest possible arrangement is the "L unit" which consists of two reactances jA and jB ohms. If R_A is greater than R_0 , say $R_A = nR_0$, $n > 1$, the arrangement is as shown in fig. 47a, whereas if $R_0 > R_A$, say $R_0 = nR_A$, the arrangement will be as shown in fig. 47b. Taking the former case, the input impedance will be

$$Z_i = jA + \frac{jBR_A}{R_A + jB} \quad \dots \quad (40a)$$

$$= jA + \frac{jBR_A(R_A - jB)}{R_A^2 + B^2} \quad \dots \quad (40b)$$

$$= \frac{B^2 R_A}{R_A^2 + B^2} + j \left(\frac{BR_A^2}{R_A^2 + B^2} + A \right) \quad \dots \quad (40c)$$

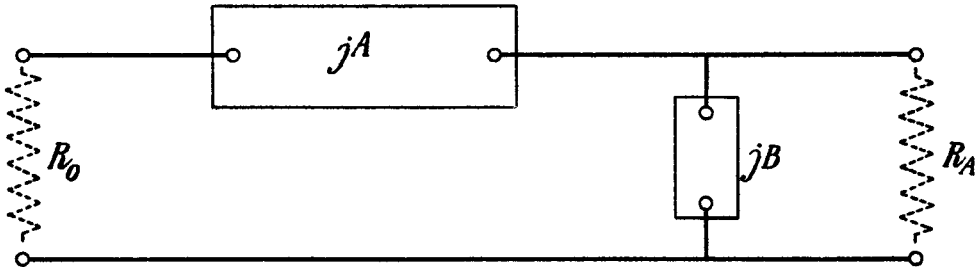
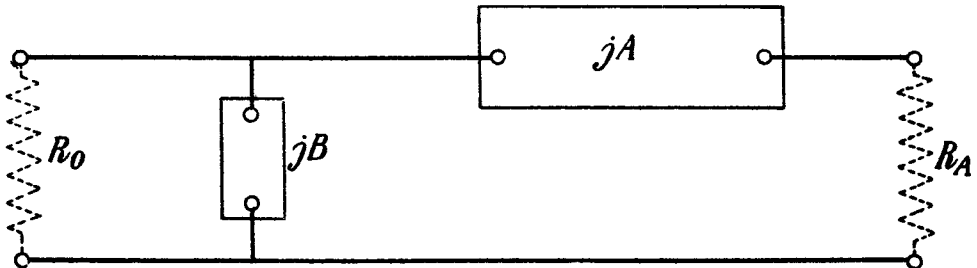

 (a) $R_A > R_0$

 (b) $R_A < R_0$

FIG. 47, CHAP. XV.—L-type matching units.

CHAPTER XV.—PARA. 110

We require Z_1 to be equal to R_0 , i.e. purely resistive, and the imaginary part must vanish, i.e.

$$A + \frac{BR_A^2}{R_A^2 + B^2} = 0 \quad \dots \dots \dots (40d)$$

$$R_0 = \frac{B^2 R_A}{R_A^2 + B^2} \quad \dots \dots \dots (40e)$$

from the above equations

$$R_0 (R_A^2 + B^2) = B^2 R_A$$

$$B^2 (R_A - R_0) = R_0 R_A^2$$

$$B^2 = \frac{R_0 R_A^2}{R_A - R_0} = \frac{n^2 R_0^3}{(n-1) R_0}$$

$$B = \mp \frac{n}{\sqrt{n-1}} R_0 \quad \dots \dots \dots (40f)$$

because $R_A = nR_0$.

Also

$$A = - \frac{BR_A^2}{R_A^2 + B^2}$$

Inserting the above value of B ,

$$\begin{aligned} A &= \pm \frac{\frac{n}{\sqrt{n-1}} R_0 \times n^2 R_0^2}{n^2 R_0^2 + \frac{n^2}{n-1} R_0^2} \\ &= \pm \sqrt{n-1} R_0. \quad \dots \dots \dots (40g) \end{aligned}$$

Thus the series and shunt reactances must be of opposite sign; if A is inductive B must be capacitive and *vice versa*. It will also be observed that $AB = -nR_0^2 = -R_A R_0$

Example.—If $R_A = 3,000$ ohms, and $R_0 = 600$ ohms, $n = 5$, $\sqrt{n-1} = 2$.

$$\therefore A = \pm 2R_0 = \pm 1,200 \text{ ohms.}$$

$$AB = -R_A R_0 = 3,000 \times 600$$

$$B = \frac{-3,000 \times 600}{\pm 1,200}$$

$$= \pm 1,500 \text{ ohms.}$$

If $f = 6$ Mc/s, and we decide that A is to be inductive, say ωL ohms,

$$\omega L = + 1,200$$

$$L = \frac{1,200}{2\pi \times 6 \times 10^6} \text{ henries}$$

$$= 31.8 \mu H$$

and B will be capacitive, say $-\frac{1}{\omega C}$ ohms.

$$-\frac{1}{\omega C} = -1,500$$

$$C = \frac{1}{2\pi \times 6 \times 10^6 \times 1,500} \text{ farad,}$$

$$= 16.67 \mu F$$

111. In the second case, $R_0 > R_A$ or $R_0 = nR_A$ (fig. 47b), and the input admittance will be

$$\frac{1}{Z_1} = \frac{1}{jB} + \frac{1}{R_A + jA} \quad \dots \dots \dots (41a)$$

$$= \frac{1}{jB} + \frac{R_A - jA}{R_A^2 + A^2} \quad \dots \dots \dots (41b)$$

Again, $\frac{1}{Z_1}$ must be equal to $\frac{1}{R_0}$, and its imaginary part must vanish, i.e.

$$\left. \begin{aligned} \frac{1}{R_0} &= \frac{R_A}{R_A^2 + A^2} \\ \text{and } \frac{1}{jB} - j\frac{A}{R_A^2 + A^2} &= 0 \end{aligned} \right\} \dots \dots \dots (41c)$$

Hence

$$R_0 = \frac{R_A^2 + A^2}{R_A} \quad \dots \dots \dots (41d)$$

$$A^2 = R_0 R_A - R_A^2$$

Substituting $R_A = \frac{R_0}{n}$

$$A^2 = \frac{n-1}{n^2} R_0^2$$

$$A = \pm \frac{\sqrt{n-1}}{n} R_0 \quad \dots \dots \dots (41e)$$

and

$$\begin{aligned} \frac{1}{jB} &= j\frac{A}{R_A^2 + A^2} \\ B &= -\frac{R_A^2 + A^2}{A} \end{aligned}$$

Inserting the above value of A ,

$$B = \mp \frac{R_0}{\sqrt{n-1}} \quad \dots \dots \dots (41f)$$

As before, A and B are of opposite sign and

$$AB = -R_A R_0$$

Example.—If $R_A = 120$ ohms and $R_0 = 600$ ohms, $n = 5$

$$B = \mp \frac{R_0}{2} = \mp 300 \text{ ohms}$$

$$A = \pm R_0 \times \frac{2}{5} = \pm 240 \text{ ohms.}$$

CHAPTER XV.—PARA. 112

Let $A = \omega L$, $B = -\frac{1}{\omega C}$, $\omega = 2\pi \times 6 \times 10^6$.

$$L = \frac{240}{2\pi \times 6} \mu H$$

$$= 6.37 \mu H.$$

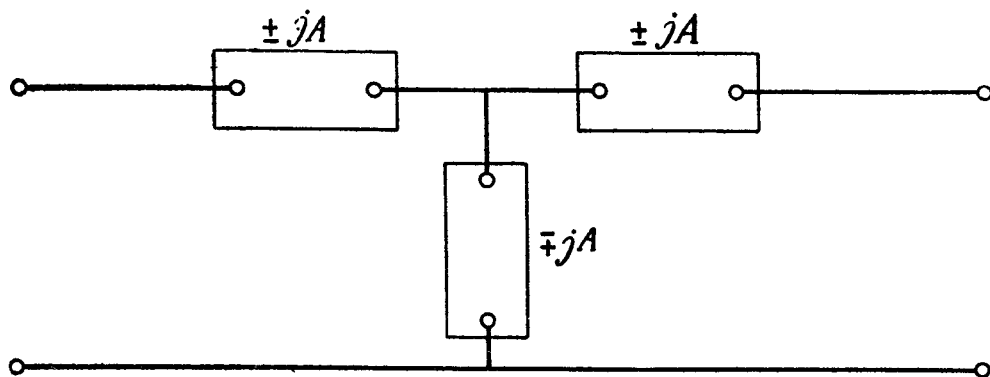
$$-\frac{1}{\omega C} = -300$$

$$C = \frac{10^9}{300 \times 2\pi \times 6} \mu\mu F$$

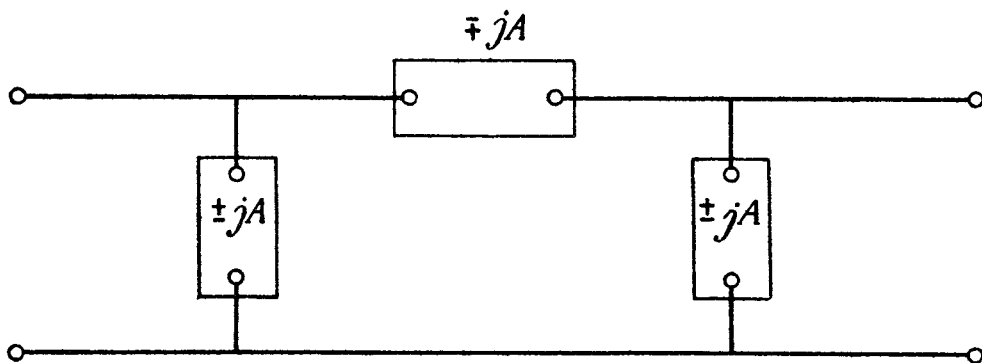
$$= 88.5 \mu\mu F.$$

Symmetrical T and Π units

112. Having shown the method of deriving the matching conditions in two of the simplest cases, we may now describe briefly certain other arrangements. Fig. 48a shows the symmetrical T, and fig. 48b the symmetrical Π networks. In each case we have three reactances each having a magnitude of A ohms. The sign of each of the series reactances is the same, but opposite to the



(a) Symmetric T unit



(b) Symmetric Π unit

FIG. 48, CHAP. XV.—Symmetric matching units.

sign of the corresponding parallel reactance. Such a network is equivalent electrically to a $\frac{\lambda}{4}$ length of transmission line of surge impedance A ohms. Hence the required reactance is immediately found by the relation $A = \sqrt{R_0 R_A}$.

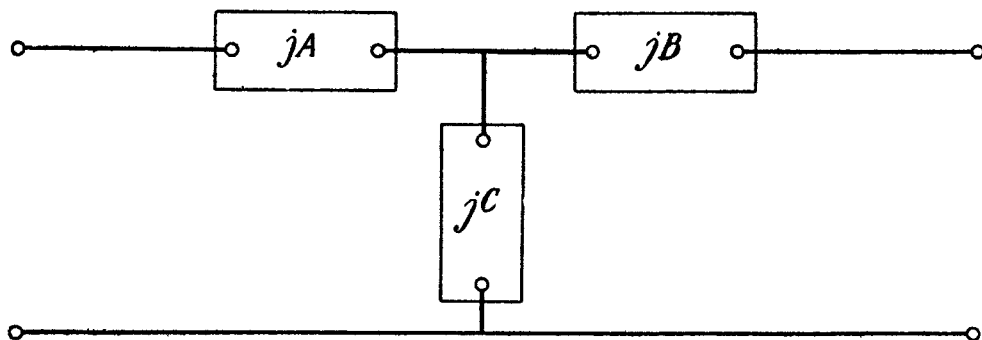
Unsymmetrical T and Π units

113. These are shown in fig. 49a and fig. 49b respectively. In each case we have three reactances A, B, C . Let $R_0 = nR_A$, where n may be greater or less than unity. Then in the T unit if $A = aR_A, B = bR_A, C = cR_A$, the numerics a, b and c are interdependent. If a suitable value is selected for C and therefore for c .

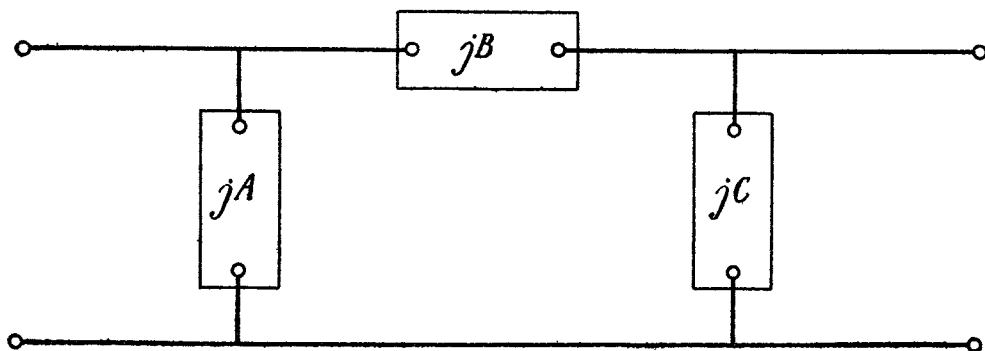
$$a = -c \pm n \sqrt{\frac{c^2}{n} - 1}$$

$$b = -c \pm \sqrt{\frac{c^2}{n} - 1}$$

In choosing the value for c , therefore, it is essential to make it greater than $\sqrt{n}$. When $c = \sqrt{n}$ the circuit becomes symmetrical. In the unsymmetrical Π unit, we have $A = aR_A, B = bR_A$,



(a) Unsymmetric T unit



(b) Unsymmetric Π unit

FIG. 49, CHAP. XV.—Unsymmetric matching units.

CHAPTER XV.—PARAS. 114–115

$C = cR_A$ as before. If the value of B is suitably selected, then, the values of a and c are related to b as follows.

$$a = -\frac{bn}{n \pm \sqrt{n-b^2}}$$

$$c = -\frac{b}{1 \pm \sqrt{n-b^2}},$$

i.e. if $b = \sqrt{n}$ the network becomes symmetrical. Hence b must be less than $\sqrt{n}$.

T — Π network

114. This is shown in fig. 50. It possesses the following important property. If the reactances A, B, C, D are so chosen that $A + B + C = 0$, and $\frac{D}{A} = \frac{B}{C} = -n$, so that D is of opposite sign to A , and B to C , the network is equivalent to an ideal transformer of turns ratio n . Thus if an

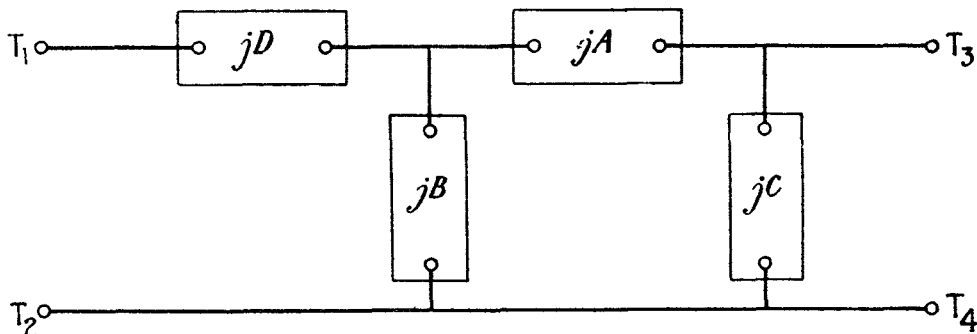


FIG. 50, CHAP. XV.—T — Π network.

aerial of resistance R_A is connected to the terminals T_3, T_4 , the input impedance at T_1, T_2 is $n^2 R_A$. To match the aerial to the line therefore, we must make $n^2 R_A = R_0$, or $n = \sqrt{\frac{R_0}{R_A}}$.

Annulment of capacitive reactance of load

115. If an aerial is connected to a feeder line in such a manner that a current node exists at a point in the aerial within a distance of $\frac{\lambda}{4}$ from the junction of line and aerial, the latter offers capacitive reactance as well as resistance and can be represented by an impedance $R_A + \frac{1}{j\omega C}$ ohms, or by an admittance $G + jB_c$ ohms, where

$$G = \frac{\omega^2 C^2 R_A}{1 + \omega^2 C^2 R_A^2}$$

$$B_c = \frac{\omega C}{1 + \omega^2 C^2 R_A^2}$$

$$= \omega C',$$

where C' is the effective shunt capacitance. Thus we may annul the reactance of the aerial by connecting, in series, an inductance $L = \frac{1}{\omega^2 C'}$, or, in parallel, an inductive susceptance $\frac{1}{\omega L'}$, where

$$\frac{1}{\omega L'} = \omega C' = \frac{\omega C}{1 + \omega^2 C^2 R_A^2}$$

or

$$L' = \frac{1 + \omega^2 C^2 R_A^2}{\omega^2 C}$$

$$= \frac{1}{\omega^2 C} + C R_A^2$$

116. If the matching is performed by means of an L, Π or T network, there is no necessity to add a physical inductance in this manner, for the actual capacitance C may be considered to form part of the matching network, the constants of the latter being adjusted accordingly.

Example.—Suppose the aerial to be terminated at a point such that at 6 Mc/s its impedance is $100 - j 25$ ohms.

$$\frac{1}{\omega C} = 25$$

$$C = \frac{1}{2\pi \times 6 \times 10^6 \times 25} \text{ farad}$$

$$= 106 \mu\text{F}.$$

To annul this we may use a series inductance

$$L = \frac{25}{2\pi \times 6 \times 10^6} \text{ henries}$$

$$= 0.66 \mu\text{H}.$$

The equivalent shunt capacitance is $\frac{C}{1 + \omega^2 C^2 R_A^2} = C'$

$$C' = \frac{106 \times 10^{-12}}{1 + (2\pi \times 10^6 \times 106 \times 10^{-12} \times 100)^2}$$

$$= \frac{106}{1.16}$$

$$= 91.5 \mu\text{F}.$$

and this may be annulled by a shunt inductance

$$L' = \frac{1.16}{(2\pi \times 6 \times 10^6)^2 \times 106 \times 10^{-12}}$$

$$= 7 \mu\text{H}.$$

Annulment of inductive reactance of load

117. If the aerial is so connected that a current loop exists in it, within a distance of $\frac{\lambda}{4}$ from the feeding point, the aerial offers inductive reactance and its impedance is $R_A + j\omega L$ ohms. Its admittance is $G - jB_L$ where

$$G = \frac{R_A}{R_A^2 + \omega^2 L^2}$$

$$B_L = \frac{\omega L}{R_A^2 + \omega^2 L^2}.$$

CHAPTER XV.—PARA. 118

Thus by connecting a capacitance C in series with the aerial, its inductive reactance may be annulled. The value of C is obviously $\frac{1}{\omega^2 L}$. Alternatively a capacitance C' may be connected in parallel, its value being given by

$$C' = \frac{L}{R_A^2 + \omega^2 L^2} \text{ farads.}$$

Again, instead of adding a physical component to the aerial itself it is possible to insert the required reactance in the last member of the matching network. Thus, no matter at what point an aerial is terminated it is always possible to ensure that its input impedance is purely resistive and suitable for matching to a non-reactive line.

Quarter-wave matching

118. We may now explain the theory of quarter-wave matching more thoroughly. Suppose a transmission line to have a surge impedance of Z_0 ohms and to be ultimately terminated by an impedance $Z_r = nZ_0$. It is therefore necessary to insert some matching device between the line

and the load. A section of line $\frac{\lambda}{4}$ long, of surge impedance Z_0' , terminated by an impedance Z_r ,

has an input impedance $Z_i = \frac{(Z_0')^2}{Z_r}$. If $Z_r = nZ_0$, $Z_i = \frac{(Z_0')^2}{nZ_0}$, and if $Z_0' = \sqrt{n} Z_0$,

$$Z_i = \frac{(Z_0')^2}{nZ_0} = \frac{n(Z_0)^2}{nZ_0} = Z_0.$$

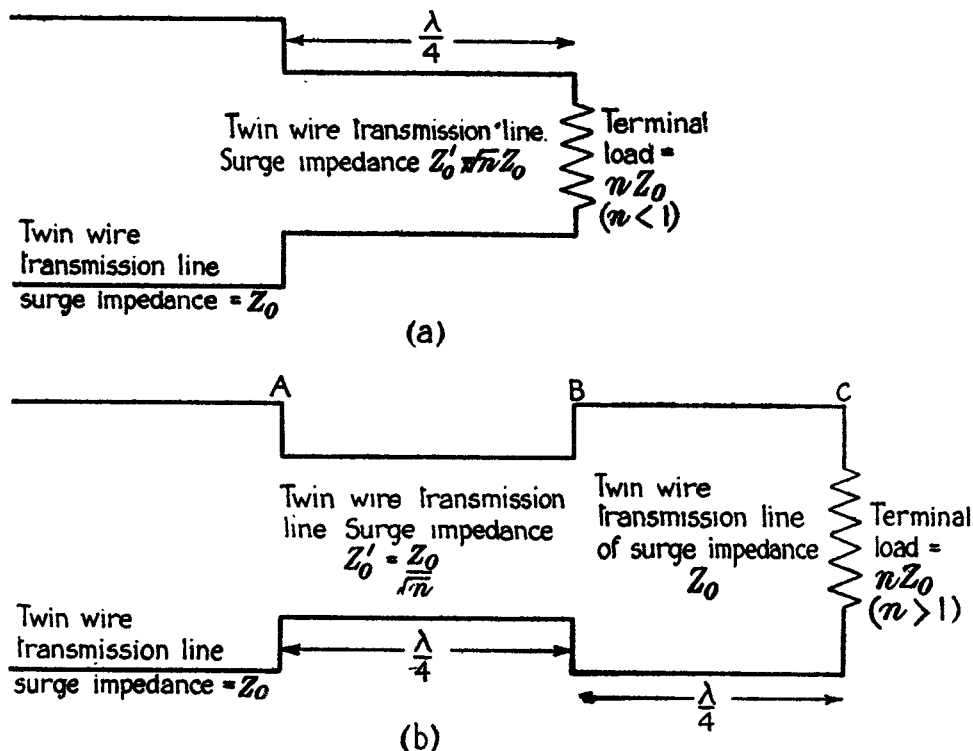


FIG. 51, CHAP. XV.—Quarter-wave matching.

Thus, if a $\frac{\lambda}{4}$ length of line, of surge impedance $Z_0' = \sqrt{n} Z_0$, is inserted between the actual line and the load proper, the line is terminated by an impedance equal to its surge impedance, which is what is required.

119. If n is less than unity, Z_0' must be less than Z_0 , and this may be achieved simply by reducing the spacing of the line over the final $\frac{\lambda}{4}$ length, as shown in fig. 51a. In effect this last section is a part of the aerial system in that it carries a stationary wave, whereas in the line proper stationary waves should be entirely suppressed. If n is greater than unity, this method of matching would entail an increase in the spacing, and consequently to increased radiation from the line in transmission, and greater pick-up in reception. It is then necessary to adopt an artifice, and arrange the feeder as in fig. 51b. The input impedance of the section B C is $\frac{Z_0^2}{n Z_0} = \frac{Z_0}{n}$, and we are back to the original problem ($n < 1$). If the surge impedance of the section A B is Z_0' the input impedance of this section is $\frac{(Z_0')^2}{Z_0}$, and we require this to be equal to Z_0 , i.e.

$$\frac{n(Z_0')^2}{Z_0} = Z_0$$

$$Z_0' = \frac{Z_0}{\sqrt{n}}$$

If these conditions are achieved, the line will be matched at the point A, and the portion A C becomes in effect a part of the aerial, carrying a $\frac{\lambda}{2}$ portion of a standing wave.

Example.—A load of (a) 120 ohms, (b) 3,000 ohms is to be matched to a 600 ohm line. Find the required conditions

(a) Here $Z_r = 120$, $= n Z_0$, $Z_0 = 600$, $n = \frac{1}{5}$. We therefore require to insert a $\frac{\lambda}{4}$ section, of surge impedance $Z_0' = \sqrt{\frac{1}{5}} Z_0 = \frac{600}{2.24} = 268$ ohms.

(b) Here $Z_r = 3,000$, $Z_0 = 600$, $n = 5$, and the line is arranged as in fig. 51b. The surge impedance of the part B C is equal to that of the line proper, namely, 600 ohms. The portion A B will, however, be of such a spacing that its surge impedance is 268 ohms.

Loop matching

120. If a pair of transmission lines is incorrectly terminated, the standing waves in the line may be as shown in fig. 52. Then at a current minimum, e.g. at A, the input impedance, looking towards Z_r is purely resistive, say r ohms, and r is less than Z_0 . At a current maximum, e.g. at B,

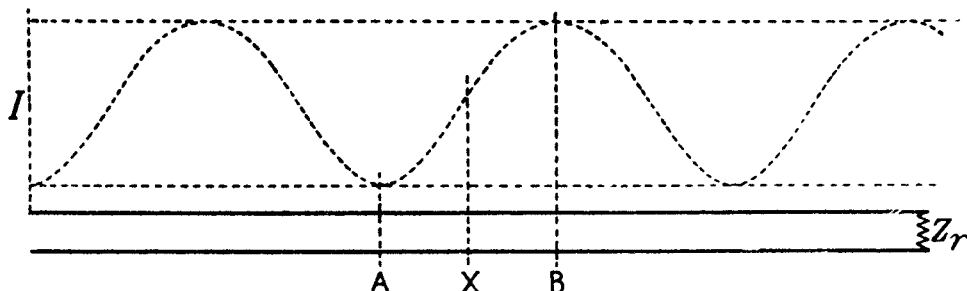


FIG. 52, CHAP. XV.—Impedance at various points in transmission line.

CHAPTER XV.—PARA. 121

the input impedance is purely resistive, say R ohms, where R is greater than Z_0 , R and r being related by the equation $rR = Z_0^2$. At any intermediate point, e.g. at X , the input admittance is complex, say $G + jB = \frac{1}{r' + jx'}$, and it is possible to match the line up to the point X (approximately) by connecting a susceptance of $-jB$ ohms so that the line becomes non-reactive at the point X .

121. It has already been shown that the input impedance of the section of line of length l , between X and Z_r , is

$$Z_x = Z_0 \frac{Z_r \cos \beta l + jZ_0 \sin \beta l}{Z_0 \cos \beta l + jZ_r \sin \beta l}$$

and its admittance is

$$\frac{1}{Z_x} = Y_x = \frac{Z_0 \cos \beta l + jZ_r \sin \beta l}{Z_0 (Z_r \cos \beta l + jZ_0 \sin \beta l)}$$

Rationalizing the denominator

$$Y_x = \frac{1}{Z_0} \frac{Z_r Z_0 (\cos^2 \beta l + \sin^2 \beta l) + j(Z_r^2 - Z_0^2) \sin \beta l \cos \beta l}{Z_r^2 \cos^2 \beta l + Z_0^2 \sin^2 \beta l} \dots \quad (42)$$

If the matching is to be achieved by the addition of a purely susceptive device, the real part of the admittance must be equal to $\frac{1}{Z_0}$,

i.e.

$$\frac{1}{Z_0} = \frac{1}{Z_0} \frac{Z_r Z_0}{Z_r^2 \cos^2 \beta l + Z_0^2 \sin^2 \beta l}$$

or

$$\frac{1}{Z_0} = \frac{Z_r}{Z_r^2 \cos^2 \beta l + Z_0^2 \sin^2 \beta l} \dots \dots \dots (43)$$

If $Z_r = nZ_0$ we have

$$\begin{aligned} \frac{1}{Z_0} &= \frac{nZ_0}{n^2 Z_0^2 \cos^2 \beta l + Z_0^2 \sin^2 \beta l} \\ \frac{1}{n} &= \frac{1}{n^2 \cos^2 \beta l + \sin^2 \beta l} \dots \dots \dots (44) \end{aligned}$$

To solve this we observe that if $\tan^2 \theta = m$, $\sin^2 \theta = m \cos^2 \theta$ and $m \sin^2 \theta + \sin^2 \theta = m^2 \cos^2 \theta + \sin^2 \theta$.

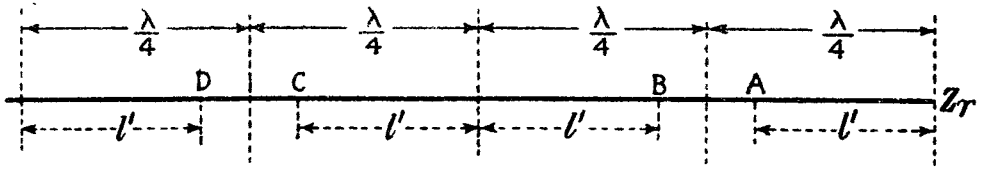
So that if $n^2 \cos^2 \beta l + \sin^2 \beta l = n$

$$n \sin^2 \beta l + \sin^2 \beta l = n$$

$$\left. \begin{aligned} \therefore \sin^2 \beta l &= \frac{n}{n+1} \\ \cos^2 \beta l &= \frac{1}{n+1} \\ \tan^2 \beta l &= n \end{aligned} \right\} \dots \dots \dots (45)$$

It follows that there is a value of βl in every quadrant of $\frac{\pi}{2}$ radians which will meet the required condition. If l' is the lowest value of l which will do this, the above equation may be written

$$l = \frac{m\lambda}{2} \pm l', \text{ where } m \text{ is a positive integer.}$$


 FIG. 53, CHAP. XV.—Physical meaning of $l = \frac{m\lambda}{2} + l'$.

This result may be translated into a physical picture of the feeder, fig. 53, in which A, B, C, D, etc., are possible positions for the matching impedance.

122. We must now consider the effect of the value of n . If n is less than unity, $\sin^2 \beta l'$ ($= \frac{n}{n+1}$) must be less than $\frac{1}{2}$. Then, $\sin \beta l'$ is less than $\cdot 707$ and $\beta l'$ less than $\frac{\pi}{4}$. Since $\beta = \frac{2\pi}{\lambda}$, $\frac{2\pi}{\lambda} l' < \frac{\pi}{4}$ means that l' must be less than $\frac{\lambda}{8}$. There will be a current maximum at the end of the line, and l' must be within the shaded areas of fig. 54a, i.e. if n is less than unity, the added susceptance must be applied within a distance of $\frac{\lambda}{8}$ from a point of maximum current.

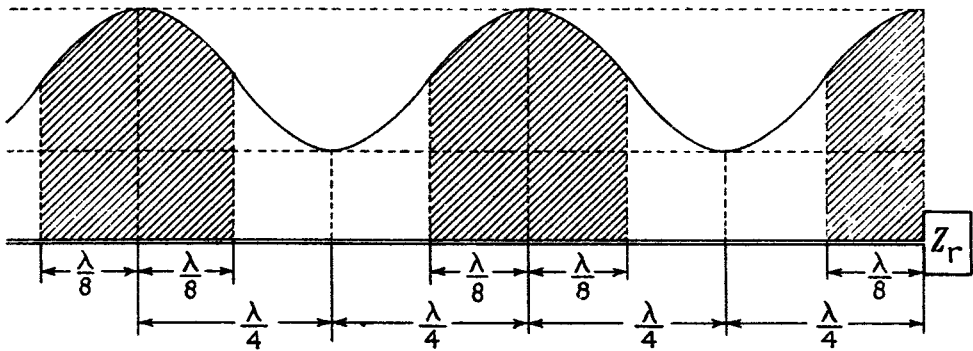
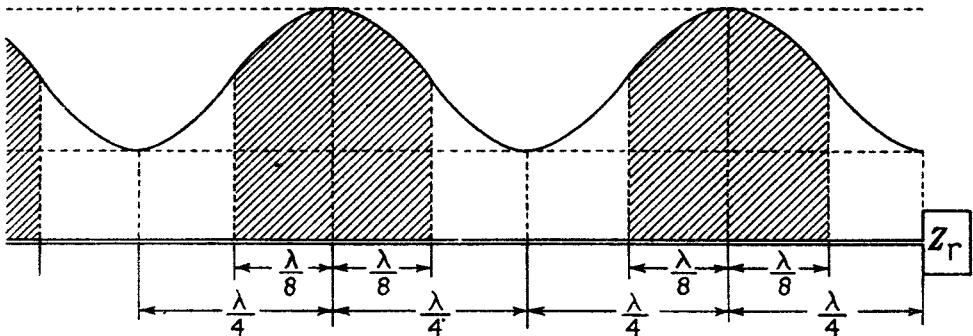

 (a) $n < 1$, current maximum at Z_r

 (b) $n > 1$, current minimum at Z_r

FIG. 54, CHAP. XV.—Positions of matching susceptance

CHAPTER XV.—PARA. 123

If n is greater than unity, $\sin \beta l'$ must be greater than $\cdot 707$ and the value of $\beta l'$ must be between $\frac{\pi}{4}$ and $\frac{\pi}{2}$. Hence l' must be between $\frac{\lambda}{8}$ and $\frac{\lambda}{4}$. The end of the line is a current minimum and l' must be within the shaded areas of fig. 54b, i.e. if n is greater than unity the matching susceptance must be applied within $\frac{\lambda}{8}$ of a current maximum as before. Thus we do not need to know whether Z_r is greater or less than Z_0 , provided that we measure l' from a current maximum. It has already been shown that the ratio $\frac{I_{\min}}{I_{\max}}$ is equal either to n or $\frac{1}{n}$, and as it does not matter whether n is greater or less than unity, the practical method is to measure $\frac{I_{\min}}{I_{\max}}$ and call this ratio n .

Substituting this value in equation (45) will give a value for $\sin^2 \beta l$ and therefore for l , fixing the possible positions at which a matching susceptance must be applied. It will be seen that there is a choice of positions. Before dealing with these, we may find the value of the matching susceptance B_m . This must be equal in magnitude but of opposite sign to the imaginary part of equation 42,

$$\begin{aligned} jB_m &= -\frac{j}{Z_0} \frac{(Z_r^2 - Z_0^2) \sin \beta l \cos \beta l}{Z_r^2 \cos^2 \beta l + Z_0^2 \sin^2 \beta l} \\ B_m &= -\frac{1}{Z_0} \frac{\left[\left(\frac{Z_r}{Z_0}\right)^2 - 1\right] \sin \beta l \cos \beta l}{\left(\frac{Z_r}{Z_0}\right)^2 \cos^2 \beta l + \sin^2 \beta l} \\ &= -\frac{1}{Z_0} \frac{(n^2 - 1) \sin \beta l \cos \beta l}{n^2 \cos^2 \beta l + \sin^2 \beta l} \\ &= \frac{1}{Z_0} \frac{(1 - n^2) \sin \beta l \cos \beta l}{n^2 \cos^2 \beta l + \sin^2 \beta l}, \quad \dots \dots \dots (46) \end{aligned}$$

since we need only consider the case when $n < 1$. Now $\sin \beta l \cos \beta l = \frac{\sin 2 \beta l}{2}$ and $\sin 2 \beta l$ may be either positive or negative. If $l = l'$, $2 \beta l$ cannot exceed $\frac{\pi}{2}$ and the sign of $\sin 2 \beta l$ will be positive. This also applies if $l = m\pi + l'$. Under these circumstances B_m will be a positive, i.e. capacitive susceptance. If, however, $l = m\pi - l'$, B_m becomes negative, corresponding to an inductive susceptance.

123. If the matching reactance is placed on the input side of a current maximum, therefore, it must be capacitive, while if placed on the opposite side it must be inductive. This is shown

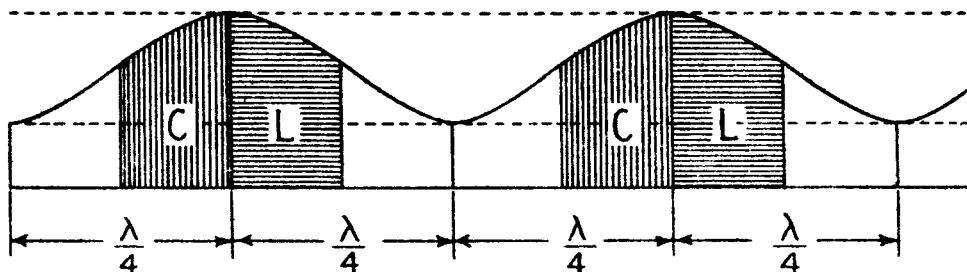


FIG. 55, CHAP. XV.—Nature of matching susceptance.

diagrammatically in fig. 55. The magnitude of the matching susceptance will now be found. From equation (46),

$$B_M = \frac{1}{Z_0} \frac{(1 - n^2) \sin \beta l \cos \beta l}{n^2 \cos^2 \beta l + \sin^2 \beta l}$$

and from equation (45), $\sin^2 \beta l = \frac{n}{n+1}$, $\cos^2 \beta l = \frac{1}{n+1}$

$$B_M = \frac{(1 - n^2) \frac{\sqrt{n}}{\sqrt{n+1}} \times \frac{1}{\sqrt{n+1}}}{Z_0 \left(n^2 \frac{1}{n+1} + \frac{n}{n+1} \right)}$$

which simplifies to

$$B_M = \frac{1 - n}{\sqrt{n} Z_0} \quad \dots \dots \dots (47)$$

For example, if $n = 0.16$, $Z_0 = 600$ ohms, $B_M = \frac{1 - 0.16}{\sqrt{0.16} \times 600} = 0.0035$ siemens (mho). At

a frequency of 2 Mc/s ($\omega = 4\pi \times 10^6$), $B_M = \omega C$ or $\frac{1}{\omega L}$.

Suppose we decide to add capacitance,

$$\begin{aligned} C &= \frac{0.0035}{\omega} = \frac{0.0035}{4\pi \times 10^6} \text{ farads} \\ &= 278 \mu\text{F}. \end{aligned}$$

If we decide to add inductance,

$$\begin{aligned} L &= \frac{1}{0.0035 \omega} \\ &= \frac{1}{0.0035 \times 4\pi \times 10^6} \text{ henries} \\ &= 22.8 \mu\text{H}. \end{aligned}$$

124. In practice the matching inductance or capacitance is usually added by means of a section of line, either on open or short circuit, as explained in paragraphs 86, 115 *et seq.* It will be found that by a judicious choice of the side of the current maximum upon which this line is connected, it is always possible for these additional "matching lines", as they are called, to be less than $\frac{\lambda}{8}$ in length. Thus, continuing the above example, we will calculate the position of the added susceptance. From equation (45)

$$\tan \beta l = \sqrt{n} = \sqrt{0.16} = 0.4$$

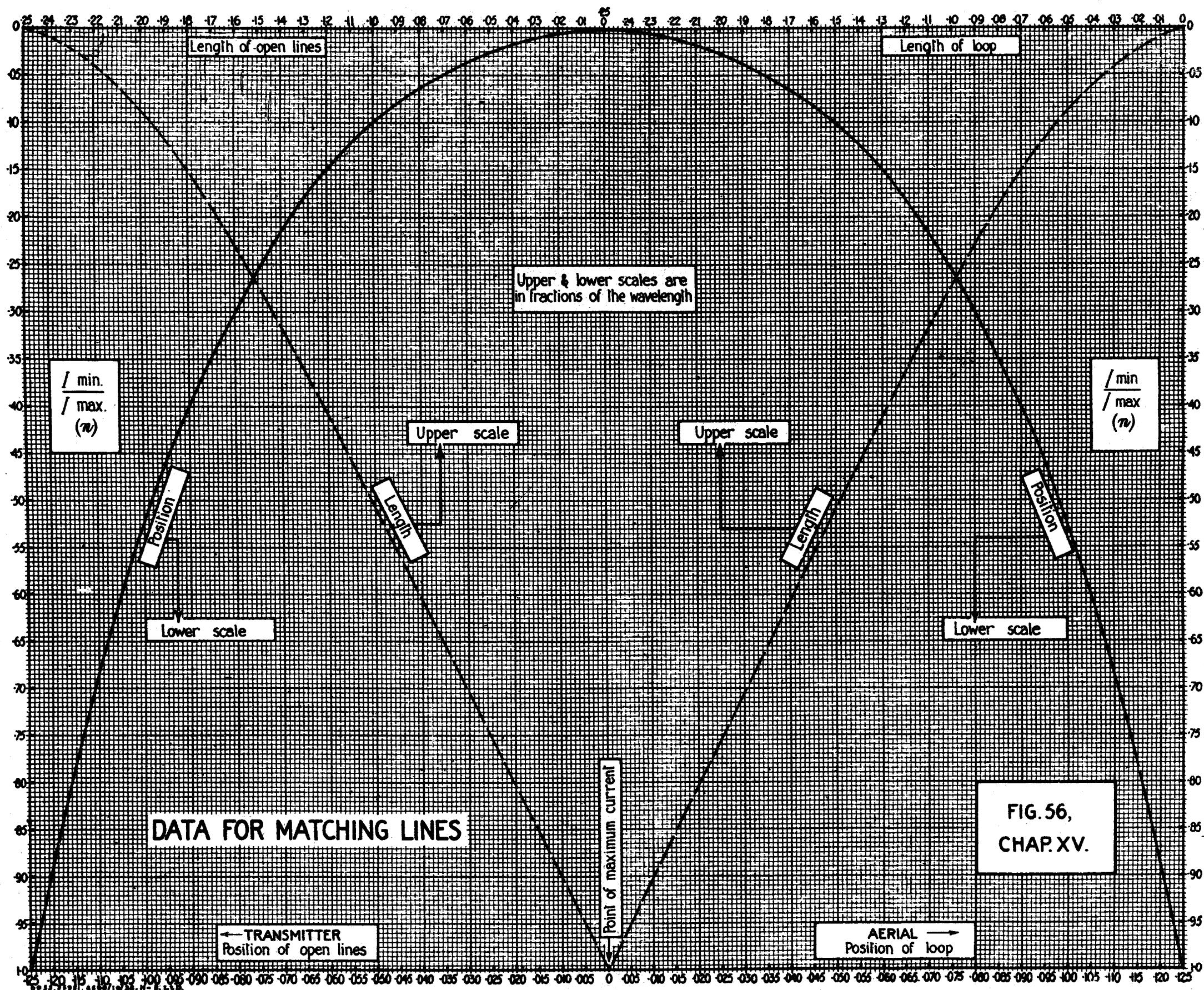
and from tables we find $0.4 = \tan^{-1} 21^\circ 48'$.

Converting to radians,

$$21.48 \text{ degrees} = \frac{21.8 \times \pi}{180} \text{ radians}$$

$$\beta l' = \frac{2\pi}{\lambda} l' = \frac{21.8\pi}{180}$$

$$\begin{aligned} l' &= \frac{21.8\pi}{180} \times \frac{\lambda}{2\pi} \\ &= 0.0605\lambda. \end{aligned}$$



CHAPTER XV.—PARA. 125

Now suppose we decide to connect the matching line on the output side of a current maximum. Then B_m must be inductive. From paragraph 86 we find that the susceptance of a short-circuited line of length l_1 is

$$B_1 = -\frac{1}{Z_0} \cot \beta l_1$$

But
$$B_m = -\frac{1-n}{Z_0 \sqrt{n}}$$

and the length l_1 must be such that

$$\cot \beta l_1 = \frac{1-n}{\sqrt{n}}$$

Continuing the example, $n = 0.16$, $\sqrt{n} = 0.4$,

$$\cot \beta l_1 = \frac{0.84}{0.4} = 2.1.$$

From tables, $\beta l_1 \doteq 25.45^\circ$

$$\begin{aligned} \frac{2\pi}{\lambda} l_1 &= \frac{25.45 \times \pi}{180} \\ l_1 &= \frac{25.45 \times \pi}{180} \times \frac{\lambda}{2\pi} \\ &= 0.071\lambda \end{aligned}$$

Note that l' gives the distance, measured from a current maximum towards the output end, at which the loop must be placed, while l_1 gives the length of the short-circuited loop of matching line which must be added.

125. Let us now find what must be done to achieve capacitive matching. The distance l' will be the same as before but must be measured from a current maximum towards the transmitter. The susceptance B_2 of a short length l_2 of open-circuited line is, by equation (31), paragraph 86,

$$B_2 = \frac{1}{Z_0} \tan \beta l_2$$

Equating as before, since $B_m = \frac{1-n}{Z_0 \sqrt{n}}$,

$$\begin{aligned} \tan \beta l_2 &= \frac{1-n}{\sqrt{n}} \\ &= 2.1 \end{aligned}$$

From tables, $\beta l_2 = 64.5^\circ$

$$\begin{aligned} \frac{2\pi}{\lambda} l_2 &= \frac{64.5 \times \pi}{180} \\ l_2 &= 0.179\lambda \end{aligned}$$

To avoid the necessity for these computations, however, fig. 56 has been developed. For any given value of n , we may read off the necessary length, either of closed (l_1) or open (l_2) matching line, from the dotted curve and top scale, and the distance l' from the current maximum by means of the full-line curve and the lower scale.

126. The foregoing theory assumes that the feeder line is terminated by a purely resistive impedance. In practice this may not be the case, but it can be shown that provided the datum point for measurement is a current maximum, the actual calculations are exactly as for a resistive termination. This is because if the line is otherwise terminated, all the current maxima and minima are shifted equally along the line.

Practical application of loop matching

127. The application of the above theory to the matching of an aerial array is as follows. The array itself will usually consist either of half-wave or quarter-wave (electrical) elements, although their actual length may not exceed $\cdot 46\lambda$ and $\cdot 23\lambda$ respectively. The transmission line may be a pair of conductors, supported upon poles as high as practicable above the ground, and clear of all irregularities of terrain. It is essential that the insulation at the points of support shall be maintained at a very high value, for a lumped leakage conductance at any point constitutes a change in the electrical character of the line and gives rise to reflection, with a consequent production of quasi-stationary waves, thus leading to both heat losses and undesired radiation from the feeder. For the same reason, the conductors should be symmetrical with respect to earth, the transmitter and the aerial array, and sharp bends must be avoided.

128. The first step in matching the array to the line is to energize the line, and observe the stationary waves in the latter. A suitable arrangement for this purpose is shown in fig. 57. It consists of a thermo-ammeter reading 0-120 milliamperes, which is mounted in a loop circuit; this loop may be suspended from one of the conductors forming the transmission line. The size of loop shown is suitable for an input into the line of the order of 1 kilowatt. The line is energized at a reduced input and the loop is drawn along it, and the current reading observed, field glasses being of assistance in this process. The current maximum at the point nearest the array is then selected for particular observation, and the power increased until the ammeter gives nearly a full scale deflection. The exact position on the line of the current maximum should be marked, and the actual scale reading, $I_{\max}$, noted. The loop is then drawn along the line to the adjacent minimum, the current, $I_{\min}$, being noted and its position marked. Especial care must be taken

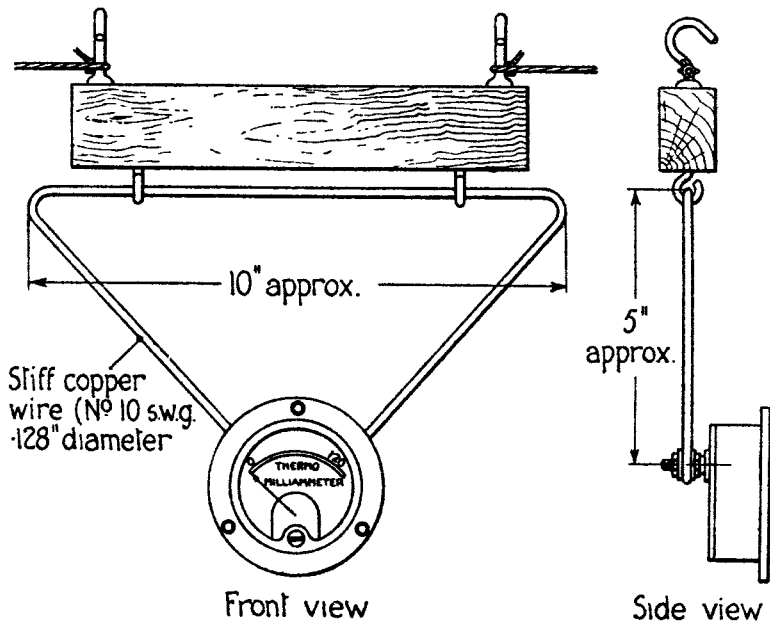


FIG. 57, CHAP. XV.—Ammeter and transformer for matching purposes.

CHAPTER XV.—PARAS. 129–131

in reading the minimum current since the lower part of the scale is very cramped. At this point it is advisable to change the ammeter over to the other conductor and verify that the currents along both lines are equal and that the maxima and minima are in the same positions in each line. If this is not so, the line and the input termination should be examined with a view to the elimination of any out-of-balance effects, as it is hopeless to attempt to match an unbalanced line. The distance between adjacent maximum and minimum positions should also be checked. This should be between $\cdot 23\lambda$ and $\cdot 25\lambda$. When all is satisfactory a final check of $I_{\max}$ and $I_{\min}$ will give the ratio $\frac{I_{\min}}{I_{\max}} = n$.

129. Referring to fig. 56, we now locate the position of the matching line by means of the curve marked "Position", e.g. if $n = 0\cdot 3$, reference to the left hand half of the diagram shows that matching may be achieved by means of an open line, distant $\cdot 080\lambda$ from the current maximum, on the side nearer to the transmitter, or by a closed loop $\cdot 080\lambda$ from the current maximum on the side nearer the aerial. Note that the bottom scale is to be read. Reference to the "Length" curve will now give the length of the matching line; in the example given, $n = 0\cdot 3$, and reading from the top scale of the diagram, we find that an open line of $\cdot 144\lambda$ or a closed loop of $\cdot 108\lambda$ will produce the desired effect, the respective positions being of course on different sides of the current maximum.

130. The choice of open line or closed loop must be governed by local circumstances. For instance if $n = 0\cdot 8$, we obtain from the closed loop curves a value of $0\cdot 116\lambda$ for the position and $0\cdot 21\lambda$ for the length, while from the open line curves we obtain $0\cdot 116\lambda$ for the position and $0\cdot 04\lambda$ for the length. If the wavelength is great compared with the height of the line above the ground, it may not be convenient to attach a loop of $0\cdot 21\lambda$ to the line, whereas an open line of $0\cdot 04\lambda$ may be only a few feet in length and easily suspended from the line. Whether open or closed matching lines are used, they must be perpendicular to the transmission line, otherwise interaction will occur between the transmission and matching lines and will give rise to losses.

131. (i) When the matching line has been attached to the transmission line the ammeter should be drawn along the latter and the maximum and minimum readings again taken; n should now approach unity. If n is less than $0\cdot 833$ ($\frac{1}{n} > 1\cdot 2$) a slight adjustment of the position of the matching line, or of its length, may improve matters, but an alteration of only an inch or so at a time should be made. If the position of the maxima and minima have interchanged, over-correction is indicated. Typical readings for an aerial consisting of a single dipole are given below.

Maximum current = $0\cdot 114$ amperes

Minimum current = $0\cdot 031$ amperes

$$n = \cdot 272$$

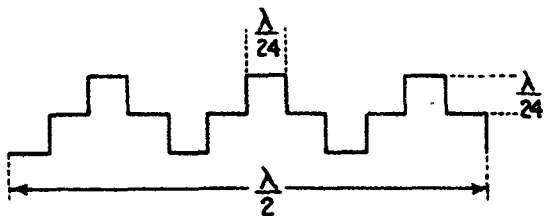
Position of matching impedance = $0\cdot 078\lambda$

Length of loop = $0\cdot 1\lambda$, or

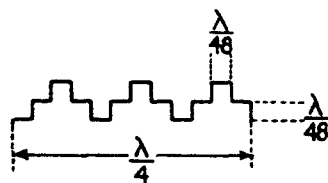
Length of open line = $0\cdot 15\lambda$

The ratio of maximum to minimum current after matching was $1\cdot 14$.

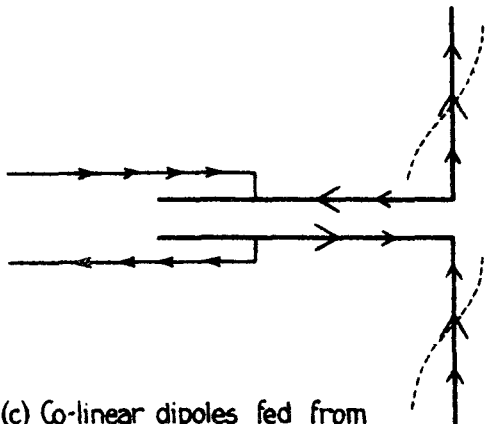
(ii) Even where matching at the aerial termination is performed by some other method, loop matching lends itself to the compensation for the lack of uniform capacitance per unit length of line. This lack of uniformity always exists to some extent because the line must in practice be supported by insulators having a permittivity greater than that of air. At frequencies of the order of 3 Mc/s the effect may be insignificant, but it will probably be appreciable at frequencies above 10 Mc/s, particularly on long lines. The symptom of such lack of uniformity is that whereas the ratio of maximum to minimum current over the first few half-wavelengths from the aerial



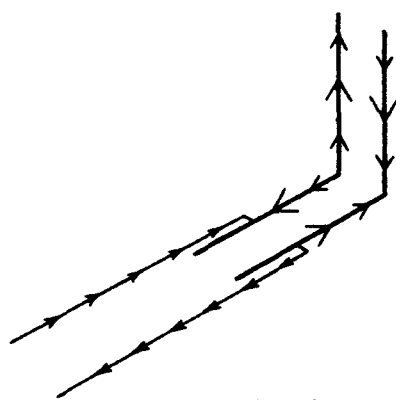
(a) One wavelength of wire acting as 1:1 transformer



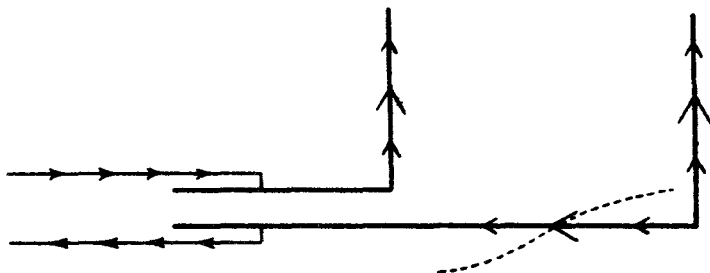
(b) One half wavelength of wire acting as -1:1 transformer



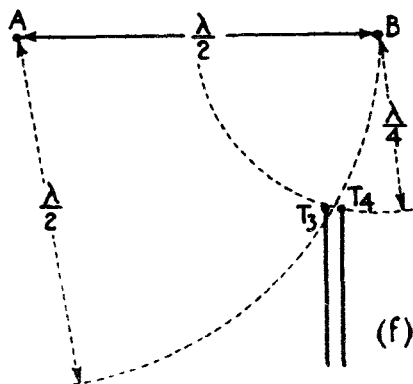
(c) Co-linear dipoles fed from transmission line.
Fields assist each other



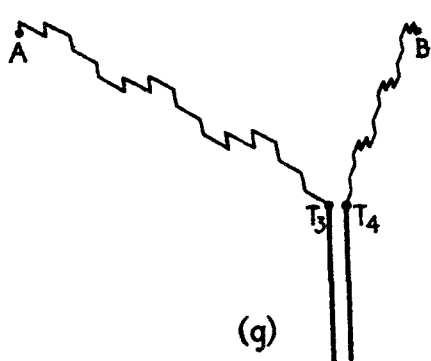
(d) Parallel dipoles fed from transmission line.
Fields cancel each other



(e) Synphased parallel dipoles fed from opposite sides of feeder line by half wave conductor acting as -1:1 transformer



(f)



(g)

end may be quite near to unity, quasi-stationary waves develop at more remote points, the mis-matching becoming more serious as the transmitter is approached. The remedy for this state of affairs is to add additional matching loops at intervals in order to maintain the $\frac{I_{min}}{I_{max}}$ ratio as near as possible to unity along the whole length of the feeder.

Suppression unit

132. It has been stated that a length λ of loss-free conductor acts as a perfect 1 : 1 ratio transformer while a loss-free length of $\frac{\lambda}{2}$ acts as a perfect transformer of ratio — 1 : 1. Although a conductor cannot be entirely loss-free, if it is arranged in such a manner as to be practically non-radiating, its loss will be almost entirely due to joulean heat and may be very small. It is possible to approach the desired non-radiating property by folding the conductor symmetrically as shown in fig. 58a and fig. 58b, in which it will be observed that the total length of wire is double the actual distance between the input and output terminals. A feeder arranged in this manner is sometimes called a "suppression unit". In fig. 58a, the feeder is required to act as a 1 : 1 transformer and the total length of wire is one λ . It is bent rectangularly at intervals of $\frac{\lambda}{24}$ forming a kind of chequer pattern in space. Similarly in fig. 58b, we have a length $\frac{\lambda}{2}$ of conductor bent into 24 rectangular loops at intervals of $\frac{\lambda}{48}$, the whole acting as a — 1 : 1 transformer. It is recommended that where this expedient is adopted the number of bent portions shall never be less than 24. The velocity of the wave along the wire is probably at least 10 per cent. less than in free space and the theoretical length of conductor should be reduced by this factor.

133. As an example of the use of such a feeder, let us consider the problem of feeding two parallel vertical dipoles, with syn-phased current from a twin-wire transmission line. If instead of being parallel, they are arranged co-linearly as in fig. 58c, they could be fed directly from the line through a suitable matching device. If however, the lower aerial of the two is turned upwards so that the dipoles are parallel instead of co-linear (fig. 58d), it is seen that the currents in the two aerials are in opposite directions, and the desired polar diagram will not be obtained. Some form of — 1 : 1 transformer must therefore be inserted in one of the aerials, and the only question is the form it shall take. If the parallel dipoles are $\frac{\lambda}{2}$ apart the most obvious method is to use a single $\frac{\lambda}{2}$ length of horizontal conductor as in fig. 58e. This conductor will however possess radiating properties and consequently the power to be supplied from one side of the line will be considerably greater than from the other, i.e. the aerial system is unbalanced. Nor is this all; unbalanced currents must flow in the transmission line and consequently this will also radiate. The final result may well be that the polar characteristics are very different from those aimed at.

134. An alternative method of feeding is shown in fig. 58f, in which the output terminals T_3, T_4 of the matching network are located as follows. From the aerial A, draw an arc of radius $\frac{\lambda}{2}$, and from B an arc of radius $\frac{\lambda}{4}$. Then T_3, T_4 are to be located at the intersection of these arcs. From the output terminal T_3 to the aerial A we may now connect a one- λ section of non-radiative

CHAPTER XV.—PARAS. 135–138

feeder as described above, which will behave as a 1 : 1 transformer, and from the output terminal T_4 to the aerial B a $\frac{\lambda}{2}$ section giving a transformation ratio of — 1 : 1. The arrangement is then as shown in fig. 58g. The two aerials will then be energized in syn-phase and will give the required polar diagram, i.e. A 5 of fig. 8. The same method may be adopted for any aerial spacing up to $\frac{3}{4}\lambda$,

for which the terminals T_3 , T_4 will lie upon the line joining the two aerials. This form of non-radiating feeder is often used to feed the vertical radiators in the Franklin uniform array, and also in conjunction with multiple unit series phase arrays.

PRACTICAL TYPES OF AERIAL ARRAY

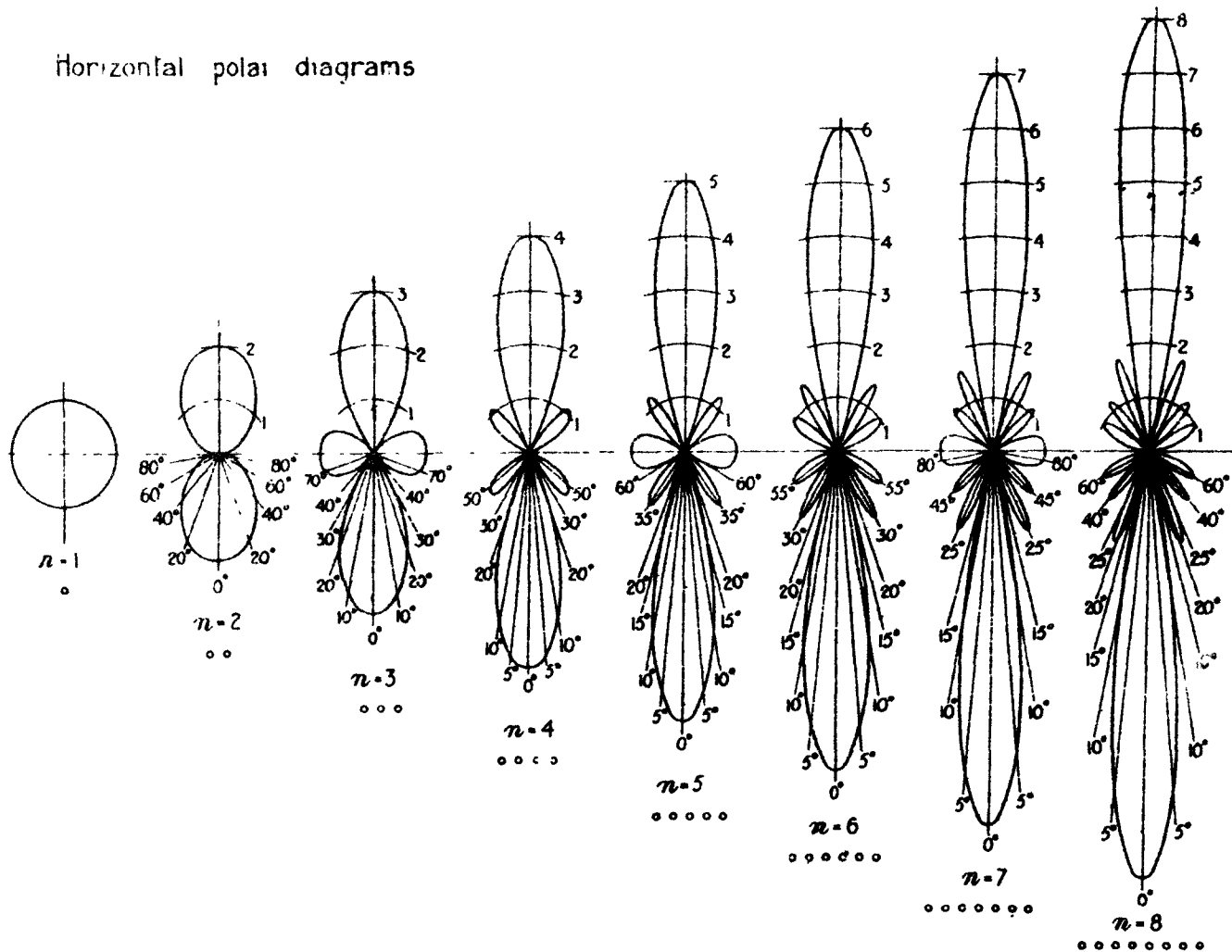
135. The simple broadside array consists of a number of aerials spaced at uniform distances along a horizontal line and fed in such a manner that all the currents are syn-phased. The width of the array, in wavelengths, is called its aperture. Fig. 59 shows the effect in the horizontal plane of an increase in the aperture of an array consisting of vertical dipoles spaced $\frac{\lambda}{2}$ apart. Each dipole is assumed to carry the same current, I , and the effect of mutual impedance between the various members is neglected. It will be observed that with n dipoles the field strength in the direction perpendicular to the line of the array is n times that given by a single dipole. To obtain this increase with a single dipole the current must increase to nI , and the power input would be proportional to $(nI)^2$. With the array of n elements, however, the power input to each is proportional to I^2 , and the total power input to nI^2 . The improvement of the array over a single dipole may be obtained from the ratio of powers and is obviously equal to n . In other words, to give a certain field strength in the required direction, an array of n elements requires only $\frac{1}{n}$ of the power which would be required by a single dipole to give the same field.

136. It must be emphasized that the improvement shown in fig. 59 can only be obtained by an appropriate increase in power supply. In order to bring out this point, fig. 60 has been prepared. This shows the horizontal polar diagrams of various arrays consisting of from one to eight elements as in the previous instance, but with the same power input in each array. The improvement is now proportional to $\sqrt{n}$ instead of to n . The effect of mutual impedance between the members may cause the improvement to be slightly less than $\sqrt{n}$, but the shape of the polar diagram is very little affected. The shape of the vertical polar diagram of such an array is given by the Current Distribution Factor for a single dipole, multiplied by the reflection diagram appropriate to the earth in the vicinity. The scale is dictated by the same considerations as that of the horizontal polar diagram.

137. The effect of a suitable reflector curtain is shown in the next diagram, fig. 61, in which each energized element is supposed to carry the same current. The effect of the reflector is to double the field strength in one of the two directions perpendicular to the array and to suppress the radiation in the other. This diagram is the theoretical one obtained with a reflector dipole placed $\frac{\lambda}{4}$ behind each energized dipole. Actually, it may be found desirable in practice either to detune the reflector wires by making them a little longer or shorter than the energized members, or to use a separation other than $\frac{\lambda}{4}$ between energized and reflector wires. Both methods may, of course, be used in conjunction.

138. In order to obtain a low angle for the main beam, the lower ends of the vertical members should be as high above the ground as possible, and in any case not less than $\frac{\lambda}{2}$. Allowing for the

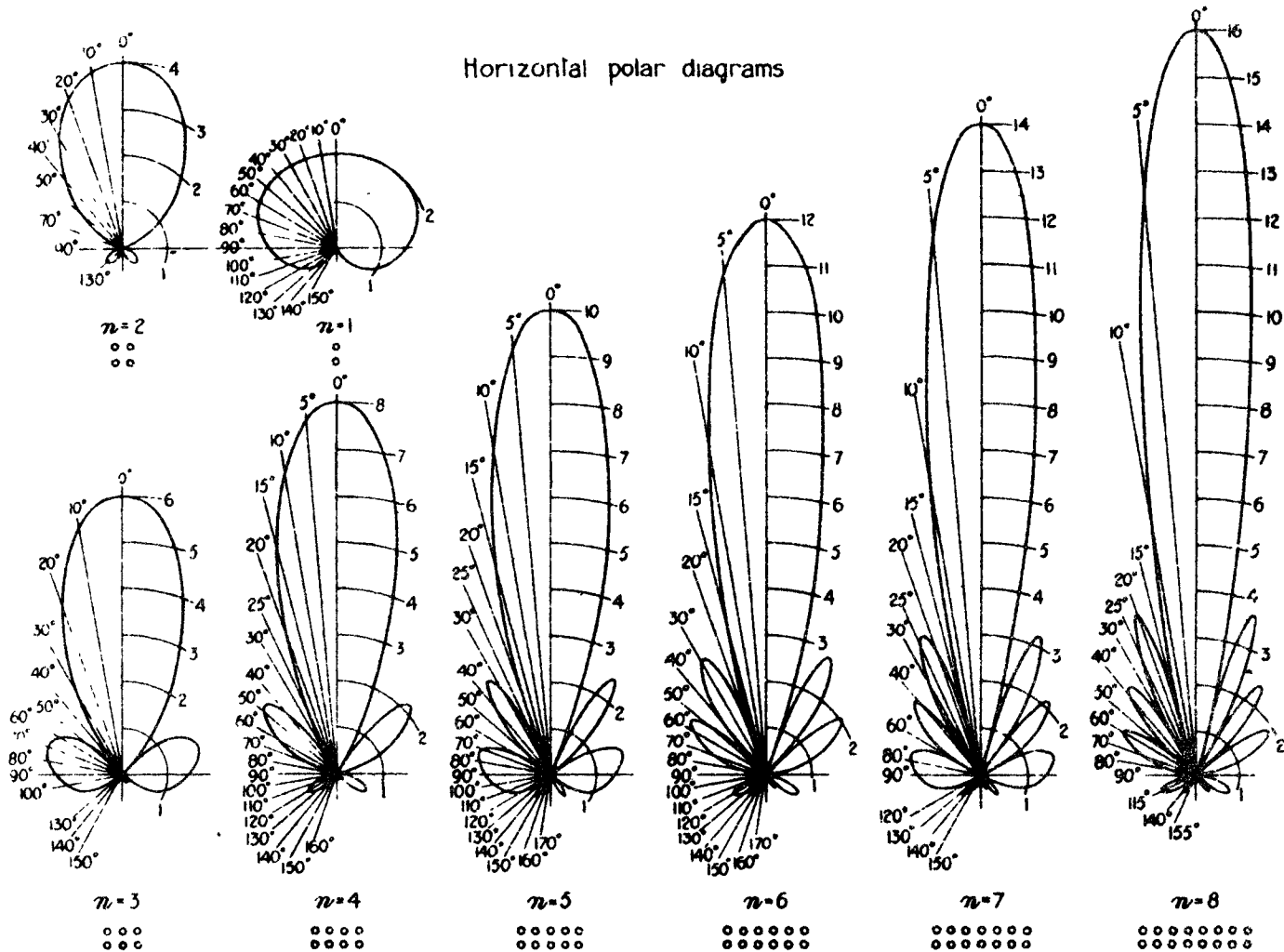
Horizontal polar diagrams



VERTICAL $\frac{\lambda}{2}$ DIPOLES, UNIT CURRENT IN EACH ELEMENT

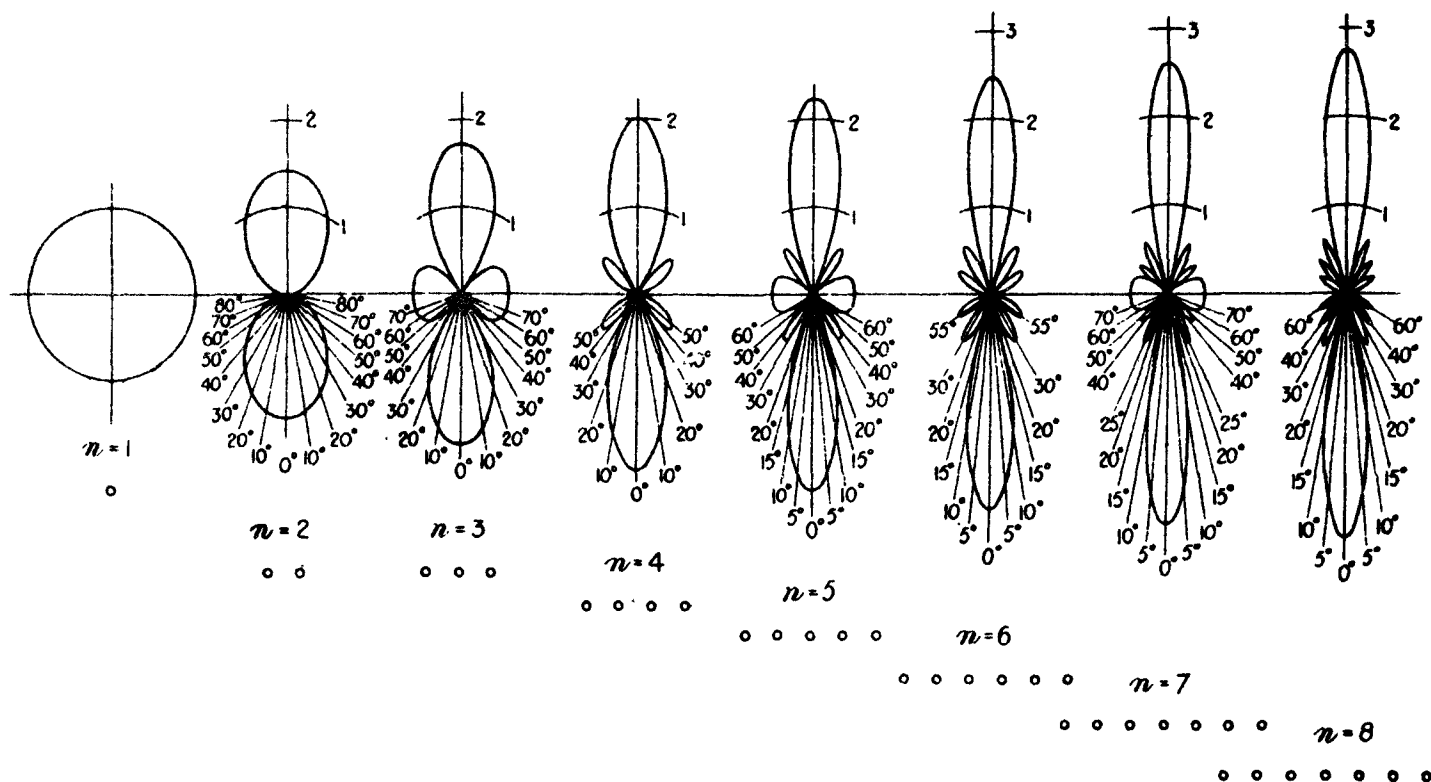
FIG. 59,
CHAP XV

Horizontal polar diagrams



VERTICAL $\frac{\lambda}{2}$ DIPOLES WITH REFLECTORS, UNIT CURRENT IN EACH ELEMENT
 FIG. 61,
 CHAP. XV

Horizontal polar diagrams



VERTICAL $\frac{\lambda}{2}$ DIPOLES , UNIT POWER IN EACH ARRAY

FIG. 60,
CHAP XV

sag in the triatic stay from which the wires are suspended, this means that the masts supporting an array of this type must be about $\frac{3}{2}\lambda$ in height. This is one of the practical disadvantages of this form of array for service purposes. Another disadvantage is the difficulty of feeding. If a feeder is attached to the lower end of each element of the energized curtain, i.e. at a voltage loop, the input impedance is of the order of 3,000 ohms, and an impedance matching device must be inserted between the feeder and the aerial element. The aerial itself is an unbalanced load, so that the alternatives presented are (i) to use a concentric feeder and some form of matching network (ii) to use parallel-wire transmission lines in conjunction with a transformer. As all the feeding points must be energized in phase, matching must be performed at a large number of points, or else a comparatively high degree of mis-matching accepted.

Tiered arrays

139. If masts of sufficient height are available, it is advantageous to arrange tiers of vertical radiators, one above the other. This results in an increase in field strength in the required direction and also gives increased directivity in the vertical plane. The problem of feeding the array also becomes somewhat simpler from the purely theoretical point of view, provided that the elements to be fed are $\frac{\lambda}{2}$ apart. It has previously been shown that a $\frac{\lambda}{2}$ length of loss-free transmission line acts as a perfect $-1:1$ transformer. It is therefore possible to arrange the feeder in the manner shown in fig. 62, alternate elements being voltage fed from opposite sides of the feeder line. The voltage distribution along the feeder is then as shown by the dotted lines, and the current distribution in the elements as shown by arrows. This method of feeding is an obvious development of fig. 58c, but in the present instance, instead of a single unbalanced conductor,

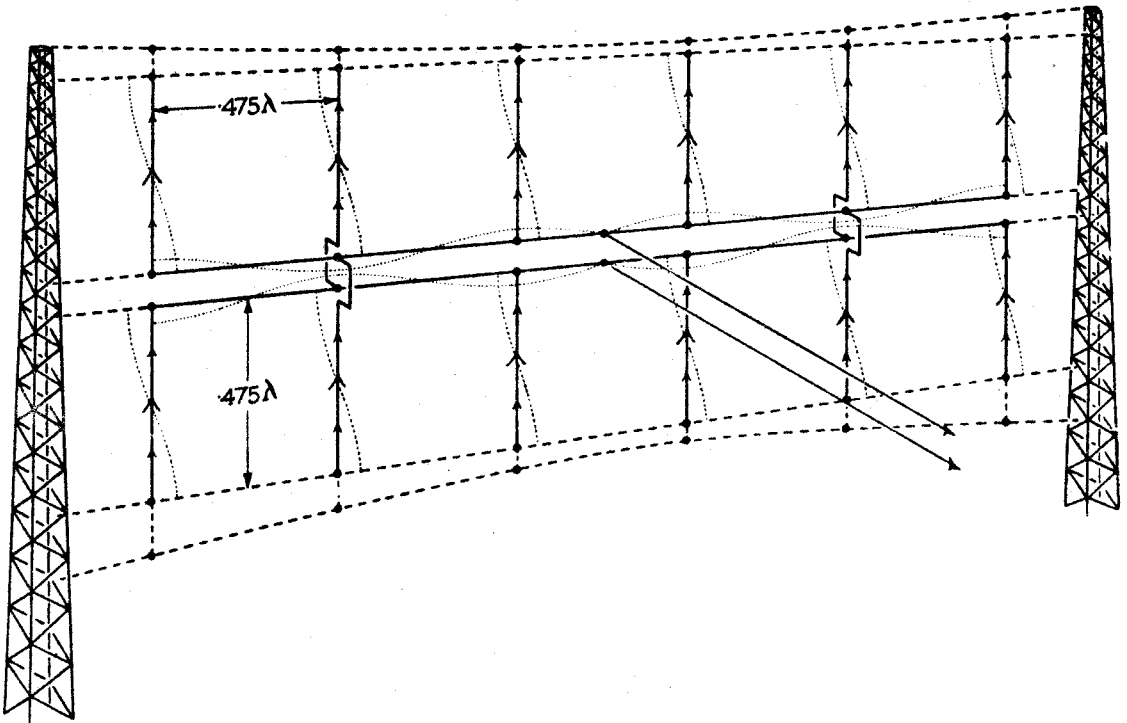


FIG. 62, CHAP. XV.—Array of vertical dipoles.

CHAPTER XV.—PARAS. 140-141

there is a twin resonant feeder line between each pair of feeding points. This feeder does not radiate appreciably and the load is very well balanced. If the bottom of the array is to be $\frac{\lambda}{2}$ above ground, the masts supporting it must have a height of about 2λ , and the feeding point will be one λ above ground. The mechanical difficulty of fitting and adjusting a matching device at this height above ground is such that the arrangement is rarely adopted.

Sterba array

140. The Sterba array is shown in fig. 63. Each unit consists of a single continuous conductor which is supplied with current at the appropriate frequency. The half-wavelength sections are arranged vertically and horizontally in an alternative manner, so that all the vertical wires carry current in phase, and set up radiation, while the horizontal sections are so arranged that

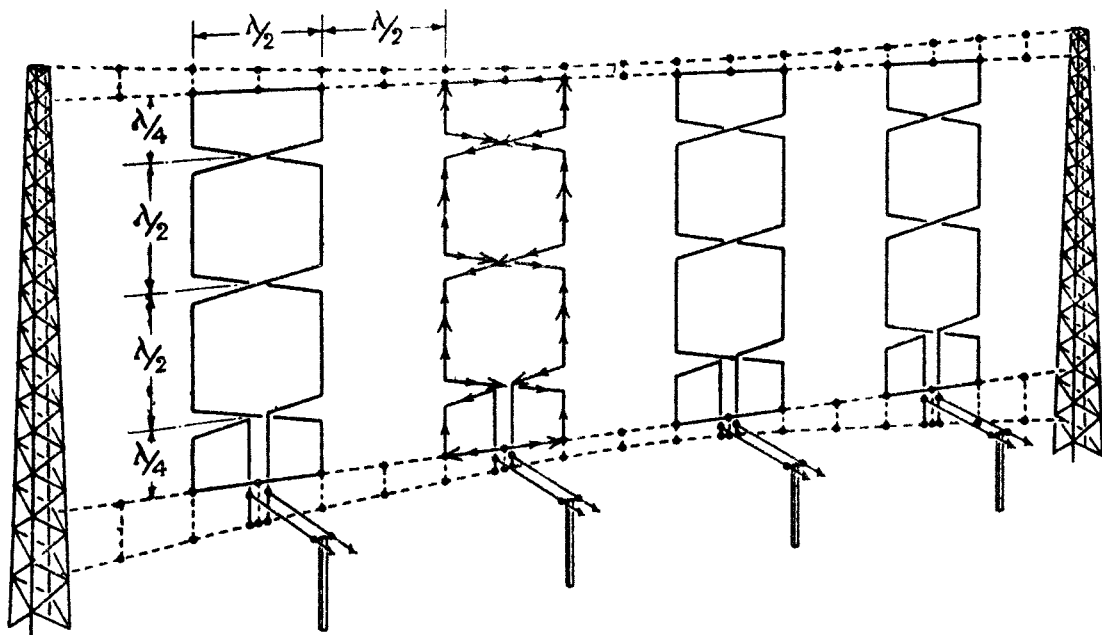


FIG. 63, CHAP. XV.—Sterba array.

they are non-radiative. The feeding points are at the current loop of one half-wave section, the unit is thus offering minimum impedance, i.e. it functions as an acceptor circuit. A number of units are erected side by side and fed with syn-phased currents by transmission lines *via* suitable matching devices. A similar array about a quarter of a wavelength behind the energized array will act as a reflector. The object of this arrangement is to allow a direct current to be fed through the radiator wires for the purpose of thawing any accumulation of snow or ice, suitable filtering devices being incorporated in the matching unit. Where "de-icing" is not necessary, the Sterba array offers no particular advantage over the simple broadside array.

End-fire array

141. An end-fire array differs from a broadside array in that there is a progressive phase difference between the currents in adjacent aerials. If β is the difference in phase and d the spacing, $\frac{d}{\lambda} = \frac{\beta}{2\pi}$. The effect of this phase progression is to cause the radiation in the horizontal

plane to be concentrated in a main lobe together with small subsidiary lobes, the main lobe being directed along the line upon which the aerials are situated. Whereas the broadside array is bi-directional, the end-fire array is unidirectional, so far as the main lobe is concerned, the radiation being directed towards the end at which the phase is lagging. The radiation in the vertical plane containing the array is more or less concentrated in a direction near the horizontal plane. The series-phase array described later is a particular example of this type.

Arrays of horizontal dipoles

142. Although the horizontal dipole gives no radiation along the earth's surface, it is found to be quite effective for long distance short-wave communication, and arrays consisting of horizontal doublets are now in extended use. They offer the theoretical advantage—which is

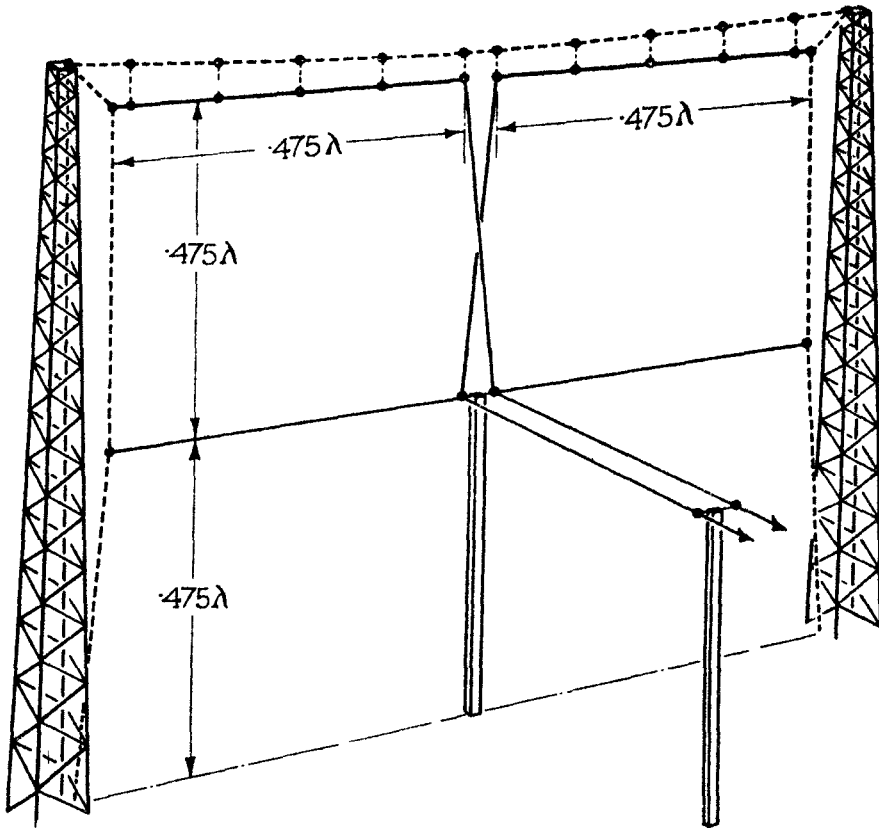


FIG. 64, CHAP. XV.—Array of horizontal dipoles.

borne out in practice—that the array is intrinsically much better balanced with respect to earth than an array of vertical elements. For the shorter wave-lengths, quite a serviceable array may be erected on 70 feet masts although, of course, higher ones are desirable for reasons already given. A very simple form consists of four $\frac{\lambda}{2}$ dipoles arranged as in fig. 64. The lower pair are connected directly to the terminals of the matching device, and the upper pair, which are $\frac{\lambda}{2}$

CHAPTER XV.—PARA. 143

above the lower, are fed by means of a $\frac{\lambda}{2}$ length of twin transmission line operating as a $-1:1$ transformer. It follows, therefore, that the feeding points must be taken from the sides of the transmission line opposite to those from which the lower side is fed. As a rough approximation, the radiation resistance of the arrangement may be taken as 4 times that of a single member less about 17 per cent. due to the effect of mutual impedance between the various members. If the lower pair are one-half wavelength above ground, the radiation resistance of each member will be about 73 ohms, and the total radiation resistance of the order of 240 ohms. A reflector curtain may be used in conjunction with the energized curtain in order to concentrate the major portion of the radiation in the required direction. It is convenient to use a reflector aerial parallel to each energized member, the spacing being $\frac{\lambda}{4}$ and the length of the reflector about 8.5 per cent.

greater than the energized member. The latter are usually 5 per cent. less than $\frac{\lambda}{2}$ so that the reflector wire has a length of about $.52\lambda$. If only 70 foot masts are available, allowing 10 feet for the sag in the triatic which supports the whole aerial, it is seen that the longest wavelength for which this aerial can be built is about 18 metres.

143. The effect of a reduction in the height of the lower members is of importance, and is very easily found to a good approximation by the use of fig. 8, and the methods explained in paragraphs 18 and 19. There is no need for extreme accuracy where only the angle and approximate magnitude of the main lobe is required. To illustrate the point, the vertical polar diagram of the four-element array has been derived, first, in fig. 65a, for the lower members $\frac{\lambda}{2}$ above earth and second, in fig. 65b, for the lower member at a height of $\frac{\lambda}{4}$. In the first diagram

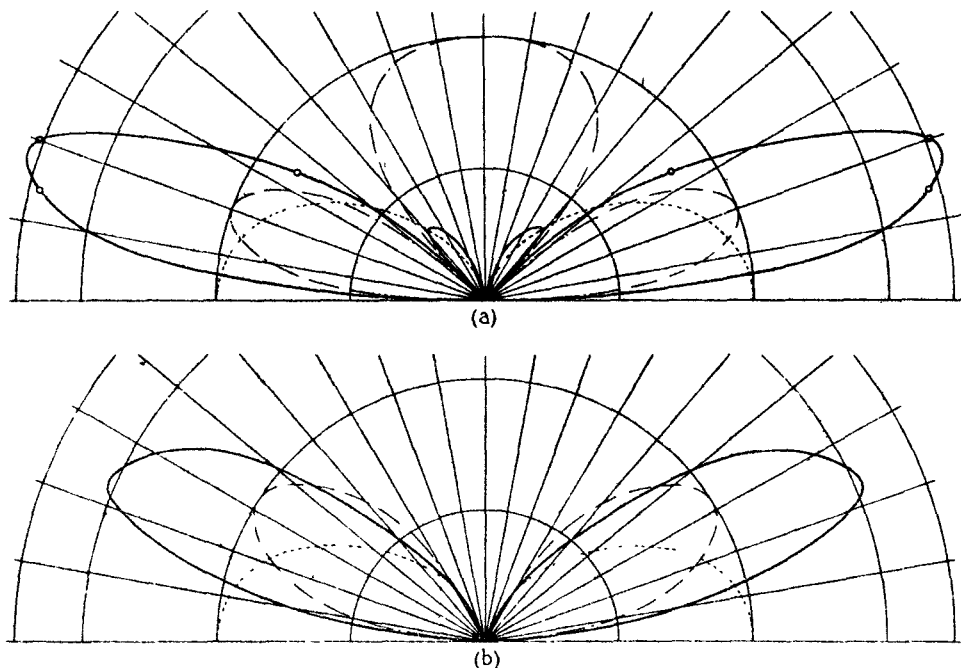


FIG. 65, CHAP. XV.—Vertical polar diagrams, arrays of horizontal dipoles.

the fine dotted-line curve corresponds to fig. 8 A 5 (parallel dipoles $\frac{\lambda}{2}$ apart with syn-phased current) and the chain-dotted curve to fig. 8 E 13 (parallel aerials $\frac{3}{2}\lambda$ apart, with currents in anti-phase) or to $h = 0.75\lambda$ in fig. 21. The product has been obtained for only four or five points and the full-line curve drawn. In the second diagram the fine dotted line is diagram A 5 of fig. 8 as before, and the chain-dotted line is obtained from fig. 8 E 9 (parallel aerials λ apart, in anti-phase) or $h = 0.5\lambda$ in fig. 21. The product is shown in full line. It is seen that in the first case the angle of the main lobe is about 18° to the horizontal, but that the field at an angle of only 4° is quite appreciable. With the lower aerial, however, the angle of the main lobe is about 22° and the field strength at angles less than 10° is very low. At a risk of over-emphasis, it is again pointed out that a few minutes study of fig. 8 and paragraphs 18 and 19, will give an approximate numerical solution of almost any example of this kind and is of greater value than many pages of purely

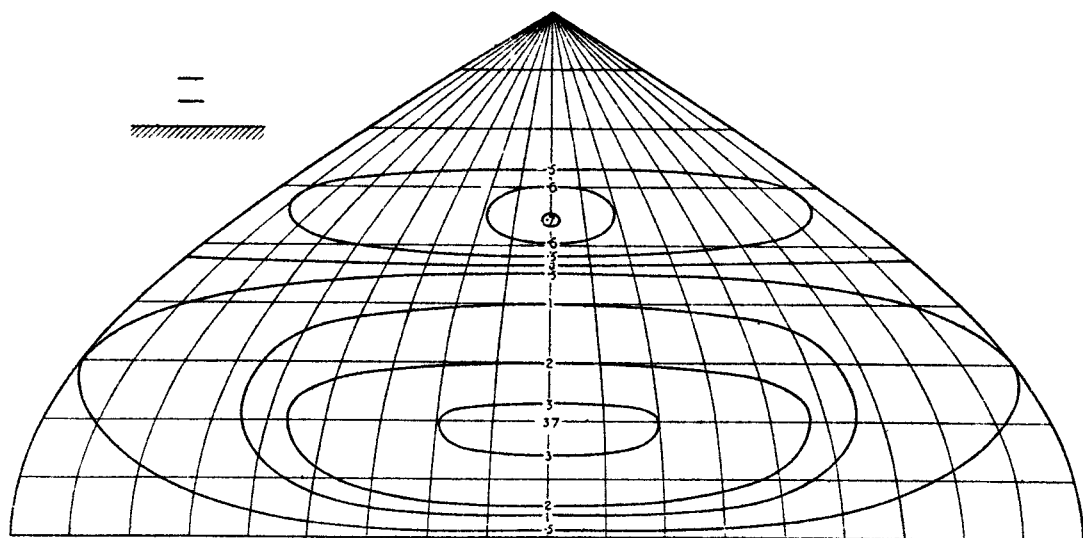


FIG. 66, CHAP. XV.—Radiation in space; array of horizontal dipoles.

qualitative statements. In fig. 66 the distribution of the field is shown upon a sinusoidal graticule for the case where the lower members are $\frac{\lambda}{2}$ above ground, corresponding with the vertical polar diagram of fig. 65a. This diagram was obtained from the latter figure by rotating it about a vertical axis through the origin, and multiplying each radius vector by the appropriate value of

the Current Distribution Factor for a $\frac{\lambda}{2}$ dipole, i.e. $\frac{\cos\left(\frac{\pi}{2} \sin \theta\right)}{\cos \theta}$, where θ is the angle in azimuth through which the vertical diagram has been rotated. The datum, $\theta = 0$, is the direction in azimuth in which the maximum radiation is produced.

The Franklin uniform array

144. This array is illustrated in fig. 67a in which the radiating members and reflecting members are shown separately. Each radiator is about 3λ in height and is doubled back upon itself in a sort of "Greek key" pattern, in order to obtain an approach to uniform current over the greater

CHAPTER XV.—PARA. 145

part of the actual height. This point is further illustrated in fig. 67c, which shows the approximate current distribution, and it will be seen that the radiation from the ends of each element of wire cancels out. As the current at these ends is comparatively small, little energy is wasted in this way, but the whole available height is made to carry a nearly uniform current approaching the loop current in magnitude. Where the mast height is sufficient, the actual radiating members are located in the higher portion of each bay, and a folded, nearly non-radiative feeder is used to convey the current from ground level to the aerial feeding points. The reflector units are usually placed about $\frac{\lambda}{4}$ or $\frac{3}{4}\lambda$ behind the radiators as shown in fig. 67b, the length of each reflector wire being adjusted to give the best forward radiation. The aperture of the array depends upon the

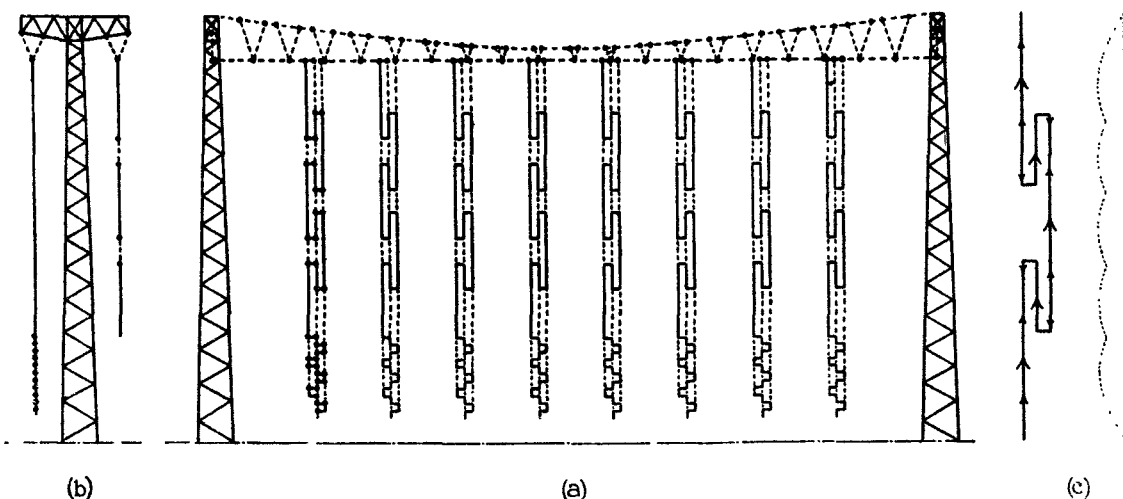


FIG. 67, CHAP. XV.—Franklin uniform array.

nature of the service ; two, four, six and eight wavelength arrays have been used in different circumstances. Although this form of aerial gives very good results, it is practically impossible to extemporize, and owing to its high cost, is now being superseded in commercial practice by the series phase array, at all events for shorter waves, i.e. 30 metres and below.

The series phase array

145. This form of end-fire array consists of a long wire, which is so bent that a series of vertical loops are formed. Each of these loops consists of a $\frac{\lambda}{2}$ section of wire doubled back upon itself, so that the height of each is $\frac{\lambda}{4}$. These loops are joined in series by horizontal portions and are separated in space by a distance of $\frac{\lambda}{4}$, the wire itself being thus continuous throughout its length.

The action will be explained with reference to fig. 68. The arrangement of the wire is shown in fig. 68a, T_1 , T_2 , being the input terminals, to which the feeder line is connected. It will be observed that T_2 is actually the earth itself, and the array is of the unbalanced type. In contrast to most of the arrangements previously described, the array may be terminated at its distant end T_0 by a non-inductive resistance equal to the effective surge impedance of the aerial, considered merely as a current-carrying conductor ; this is about 300 ohms. When so terminated, no standing

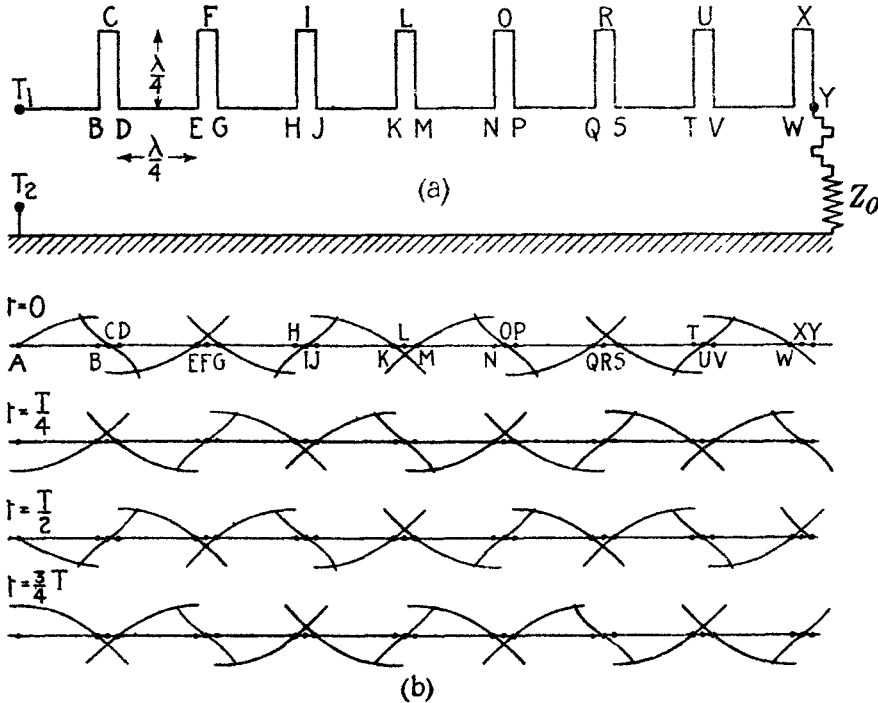


FIG. 68, CHAP. XV.—Series phase array.

waves are set up in any portion of the array—a point of primary importance. If a P.D. is applied to the terminals T_1 , T_2 , a travelling wave will be set up in the wire, moving from left to right. Considering only the vertical portions, it will be seen that at any given instant the current at all points in the section BC will be equal and opposite to that in the adjacent section CD. As these are so close together they may be considered as a single radiator carrying equal anti-phased currents, and we have seen that the effect of such currents in a $\frac{\lambda}{4}$ length of wire is to set up one quarter of a standing wave in the wire. Thus, in effect, the loop BCD acts as a $\frac{\lambda}{4}$ aerial. Similar considerations apply to the loops EFG, HIJ, etc. It must be understood, however, that in these successive loops the effective standing waves of current are not in phase with each other. The phase difference between any two successive loops will depend upon the length of horizontal connecting wire, and when this is $\frac{\lambda}{4}$, the standing wave in EFG will reach its maximum a quarter of a cycle earlier than that in BCD and so on. The current distribution at four successive intervals (of $\frac{1}{4f}$ seconds) is shown in fig. 68b. It is also seen that, at any given instant, the current in adjacent horizontal sections is in anti-phase, and consequently the total radiation from these portions is negligible. The radiation resistance R_r of a vertical $\frac{\lambda}{4}$ radiator is equal to about 36 ohms, and if I is the R.M.S. current at the base, the power radiated is $P = I^2 R_r$. Now each vertical member

CHAPTER XV.—PARAS. 146-147

of a S.P. unit acts as a $\frac{\lambda}{4}$ radiator, but its base current is effectively equal to twice the feed current. For a given feed current therefore, the power radiated is four times that which would be radiated by a $\frac{\lambda}{4}$ aerial with the same feed current, and the radiation resistance of each vertical member of a series phase array is therefore of the order of 144 ohms.

146. The horizontal polar diagram of a single-unit series phase array depends upon the number of vertical loops. If only two loops are used, the diagram approximates to fig. 8 A 3, i.e. a cardioid maximum radiation occurring toward the input end. If the length of the unit is extended with a corresponding increase in the number of verticals, the main lobe of the diagram becomes sharper, subsidiary lobes of small magnitude being developed. In practice the system is sometimes extended to a length of 4 to 6λ , i.e. from 17 to 25 verticals. When so extended, the attenuation of the current cannot be entirely neglected. It must be observed that since the radiating elements are in series, and each has a comparatively high radiation resistance, the attenuation is very much greater than in a non-radiating line of the same length. In the latter also, the whole of the power transmitted down the line is absorbed at the terminating resistance. A little reflection will show that if the attenuation is very great, the loops nearest the transmitter will radiate well, but the remote ones poorly, and the polar diagram will not be sharply directive, while if the attenuation is low, the majority of the power supplied to the array will be dissipated in the termination, and the efficiency will be very low. Thus, for transmission, there is an optimum length, which is of the order of 4λ to 5λ . Under these conditions, the ratio of currents in the first and last members may be of the order of 6 to 1, or a power ratio of 36 to 1. The terminating resistance is then practically unnecessary. It follows that the theory is more complex than was suggested above, in that, instead of a travelling wave, quasi-stationary waves will be set up in the system.

147. In the foregoing explanation, the lengths of the various vertical and horizontal elements were said to be $\frac{\lambda}{4}$. This is, of course, the electrical and not the actual length. Owing to the method of construction, in particular the large number of sharp bends, and to the effect of mutual impedance between the radiators, it is found that the verticals should be about $\cdot 225\lambda$ to $\cdot 23\lambda$ in height and the same distance apart. This is of importance in obtaining the desired "end-fire" polar diagram. Another point of practical significance is the attenuation of the current in successive radiators. If the feed current is I amperes, the ratio between the currents in successive radiators being x (< 1), and the total resistance of a unit consisting of one horizontal and one vertical element is R ohms, the total power dissipated will be

$$R \{ I^2 + (xI)^2 + (x^2I)^2 + (x^3I)^2 \dots \}$$

$$= RI^2 \{ 1 + x^2 + x^4 + x^6 : \dots \text{to } n \text{ terms} \}$$

the final term of the series representing the power dissipated in the terminating resistance. Hence the power input to the whole array is

$$\frac{1 - x^{2n}}{1 - x^2} I^2 R \text{ watts,}$$

and the input resistance $\frac{1 - x^{2n}}{1 - x^2} R$ ohms. In practice x^{2n} is very much smaller than unity

and the input resistance approaches the value $\frac{R}{1 - x^2}$ ohms. For instance, if $x = \cdot 8$ the input

resistance will be $\frac{R}{1 - 0.64} = 2.78 R$. As R may be about 160 ohms, the input resistance is about

450 ohms for this particular attenuation. This calculation again ignores the effect of mutual impedance, which causes the radiation resistance of each successive member to differ from that of the preceding one.

148. The array was originally suspended with its horizontal members at a height of $\frac{\lambda}{4}$ above ground, but better results appear to be obtainable if this height is increased. The nature of the soil under the array is also of importance. Best results appear to be obtained when the ground is either very highly conductive or almost perfectly insulating but of low permittivity, and the moist earth of the average site in Britain appears bad. There are, however, little data yet available in these respects. The frequency toleration of the series phase array is only of the order of 2 per cent. This constitutes a considerable disadvantage for service purposes. The directivity of a series phase array consisting of 8 loops is shown in fig. 69a, and the vertical diagram in fig. 69b. The horizontal directivity can be improved by using two parallel arrays fed in

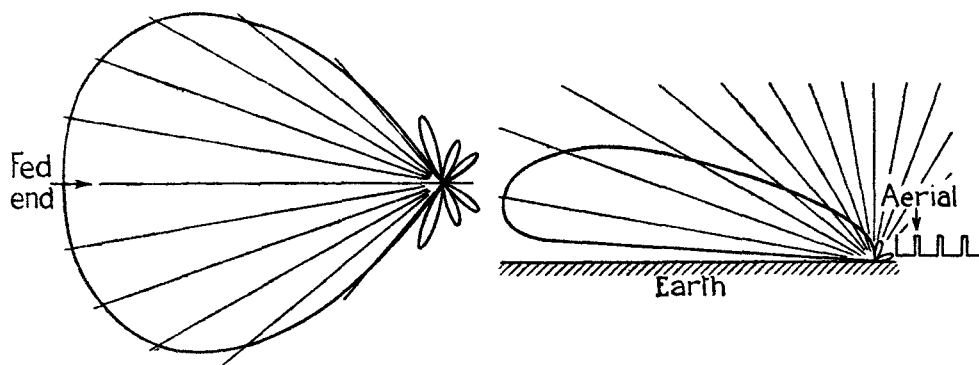


FIG. 69, CHAP. XV.—Polar diagrams of 2λ series phase array.

syn-phase. These may be $\frac{3}{4}\lambda$ apart, for convenience in feeding by means of a non-radiative feeder as described in paragraph 134. Two parallel arrays may then be connected, *via* a suitable matching device, to a twin wire feeder, and will constitute a balanced load.

Arrays used for reception

149. In general, any of the forms of array which have been previously described may be used for reception, the directional properties being practically the same for either purpose. Since, however, arrays are generally used where the traffic is continuous, it is rarely required to use a given array for alternate periods of transmission and reception; in any case, the transmitter is usually remotely controlled and it is most convenient to erect an array for the sole purpose of reception. It is then obviously uneconomic to adopt arrangements which may be imposed by transmitting considerations, e.g. breakdown voltage does not enter into the design of a receiving array. On the other hand, correct termination is just as important if not more so than in the transmission case, and due attention must be paid to the nature of the aerial, balanced or unbalanced as the case may be, in designing the matching units. A single dipole opened at its centre for the connection of the feeders has a total resistance of the order of 100 ohms, and may be directly connected to a feeder of $Z_0 = 100$ ohms. Where the length of line is not too long a length of ordinary twisted flex may be used as a feeder, for its surge impedance is of this order. This arrangement is also suitable for transmission when the input does not exceed a few watts. As the insulating material between the conductors is partly air and partly of cotton, rubber, etc., the losses are rather greater than in an open line.

150. The series phase array is finding increasing favour for receiving purposes. Since the field is not uniform in phase over the whole of the array, it is rarely advantageous to extend the length beyond about $2\frac{1}{2}\lambda$, i.e. eleven vertical loops. In some cases the feeder end is elevated above the remote end in order to obtain additional vertical directivity. If two parallel arrays are used they

CHAPTER XV.—PARA. 151

may be spaced $\frac{3}{4}\lambda$ apart and connected *via* a suitable matching device to each side of a twin wire feeder line through λ and $\frac{\lambda}{2}$ suppression units respectively. At the receiving end the received currents are then in the correct phase for connection to the input terminals of a balanced receiver, the line being also balanced and therefore practically non-radiative. The signal-noise ratio of this arrangement is found to be of a high order compared with that of a single dipole.

Rhombic array

151. This type of array has several useful forms, e.g. a single tilted wire, an inverted V, or a horizontal diamond shape. These are all classed together because the same principles are involved. The precise form adopted in any given case depends upon the polarization of the incoming wave, the direction, the wave tilt, the frequency, the available space and the material available for construction. The original form was the tilted wire aerial shown in fig. 70. First, suppose the wire

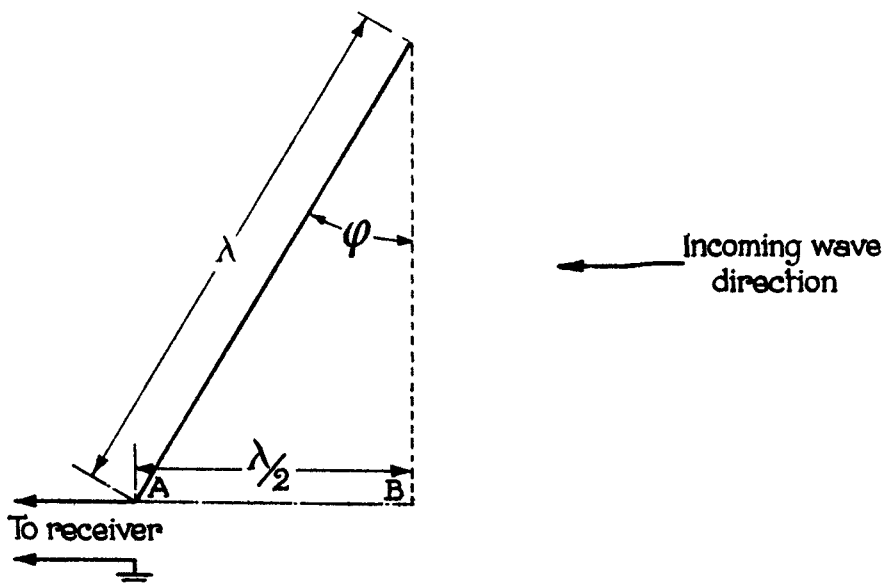


FIG. 70, CHAP. XV.—Tilted wire aerial.

to be several quarter-waves in length, erected vertically and connected to earth by a non-reflective terminating impedance. If the electric field vector E of an incoming wave is vertical, it will on arrival at the aerial induce in any element of length l an E.M.F. lE , and a voltage wave due to this will travel both upward and downward. The former wave will be reflected at the free end and will travel downward to the termination, so that in effect, there is chiefly a voltage wave downward. Since every elementary wave of strength l originates at a different point in the wire, they do not arrive at the termination in phase. The current in the terminating resistance is therefore due to the resultant of a number of elementary voltage vectors, and the magnitude of this re-

sultant depends upon the length of the wire. If the length is $\frac{\lambda}{2}$, the resultant is a maximum, and

if it is a whole wavelength the resultant is zero. If however this one-wavelength wire is tilted forward in the direction of the transmitter, any given phase of the electric field vector reaches the upper portion of the wire before the lower, and consequently the induced E.M.F.'s in the upper portion are advanced in phase with respect to the lower; this phase advancement is obviously

progressive as we consider elements further from the termination. If then the tilt is such that the upper end is $\frac{\lambda}{2}$ nearer to the transmitter than the lower end, the current vectors due to the various voltage elements will all be in phase.

152. The angle which the tilted wire makes with the horizontal is thus very important ; maximum energy is delivered to the receiver when the base AB of the triangle formed by the wire and the ground is $\frac{\lambda}{2}$ less than the length of wire. For a length l , $AB = l - \frac{\lambda}{2}$, e.g. if $l = \lambda$, the tilt angle φ , measured between the wire and the vertical, is $\sin^{-1} 0.5$ or 30° . If $l = 2\lambda$, $AB = \frac{3}{2}\lambda$ $\varphi = \sin^{-1} \frac{1.5\lambda}{2\lambda} = 49^\circ$ and so on. As a result of this relation, the optimum tilt angle varies only very slowly, if the length is greater than about 4λ , and consequently the same aerial is effective over a fairly wide frequency range.

153. In practice, the above form is rarely used because it is possible to obtain better results without an appreciable increase in material. The simplest development is to place two similar tilted wires back to back forming an inverted V. If one end of this is open and the other connected to the receiver (the latter being properly matched to the aerial) the aerial has a broadly bi-directional response, but if the free end is earthed through a terminating resistance equal to the surge impedance of the aerial, its response becomes practically unidirectional, receiving mainly from the direction in which the termination is situated (fig. 71a). The forms shown in figs. 70 and

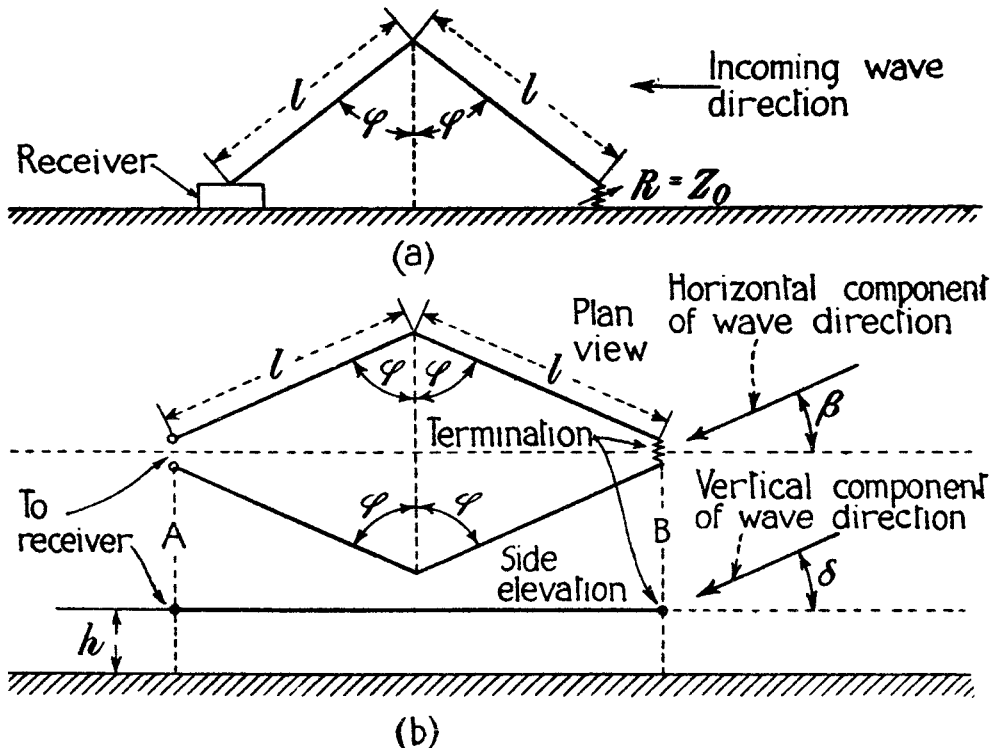


FIG. 71, CHAP. XV.—Inverted V and horizontal diamond arrays.

CHAPTER XV.—PARAS. 154-155

71a are intended for the reception of vertically polarized waves, but experimental results showed that the horizontally polarized component of down-coming waves would provide ample field strength for long distance reception, while an aerial suitable for such reception would have comparatively little pick-up of vertically polarized waves and might therefore be expected to give a high signal-noise ratio. As a result the horizontal diamond array was evolved. This is shown in plan and elevation in fig. 71b. In its simplest form it consists of two horizontally opposed V sections, similar to fig. 71a, one end of the array being connected to the receiver, and the other terminated by a suitable resistance. As in the tilted wire type, the length of each element decides the maximum gain of the array. The latter is obtained by making l exceed by $\frac{\lambda}{2}$ the length of its projection upon the base A B.

154. In the design of a horizontal diamond array the three variables to be adjusted are l , φ and h (fig. 71b). The angles β and δ are regarded as constant. Then the lowest permissible height is

$$h = \frac{\lambda}{4 \sin \delta}$$

while φ is given by

$$\sin \varphi = \cos \delta$$

and for maximum gain

$$l = \frac{\lambda}{2 \sin^2 \delta}$$

155. In practice it is found that the greater l is, the wider is the efficient reception band of a particular aerial. The optimum value of the terminating resistance, for a high front to back reception ratio, is found by trial. The actual conductors forming each inverted V are frequently constructed of twin parallel wires, connected in parallel and spaced a few inches apart. By varying the spacing it is then possible to facilitate the matching between the aerial and the transmission line feeding the receiver. It is also possible by this means to make the surge impedance of the aerial uniform at all points throughout its length, and so to decrease the power loss in the aerial. To do this the twin wires are spaced apart at the apex of each V and close up towards the opposite ends of the wires. The diamond array may be used for transmission, but little data are available as to its performance. The directivity should, on theoretical grounds, be similar to that of the same array used as a receiver.

CHAPTER XVI.—RADIO AIDS TO NAVIGATION

D/F THEORY**Introductory**

1. It is a remarkable fact that directional transmission and reception were realized in the earliest experimental work in connection with electro-magnetic radiation. The fact that the waves produced by Hertz in his classical research upon the radiation from an open oscillator could be reflected and refracted in a manner similar to light is regarded as a proof that light and hertzian waves are fundamentally of the same nature. Hertz used waves of the order of 0.5 metre in length, and was successful in producing beams of radiation by the use of parabolic reflectors, a technique which is in use to-day for certain radio-telephonic channels over comparatively short distances. The discovery (by Marconi) of the remarkable radiating properties possessed by the vertical earthed aerial wire caused a temporary suspension of research in the sphere of directional transmission and reception, although it was recognized that if transmitters and receivers could be given accurately directional properties, they would be of enormous value in marine and aerial navigation. The application of directional transmission as a means of telegraphic and telephonic communication over large distances leads to a considerable economy in capital and working costs. Certain forms of directional transmitter are also used purely for navigational purposes. They are collectively referred to as radio beacons, and will be briefly described in this chapter.

Direction finders

2. A direction-finder may be defined as a radio receiving station having an aerial system which possesses directional properties. By virtue of these it is possible to find the horizontal angle of incidence of an electro-magnetic wave arriving at the location of the receiver. A direction-finder thus gives a position line, which may in certain circumstances be made use of in obtaining a fix *i.e.* to ascertain the position of an aircraft. The words position line and fix are technical terms used in navigation, and are defined in A.P. 1234 Manual of Air Navigation, Vol. I. In general at least two position lines are necessary to obtain a fix; these may both be obtained by visual observation, one visually and one by radio, or both by radio. It may be here noted that the bearing obtained by radio is ordinarily subject to an ambiguity of 180° , as will be more fully explained later. The ambiguous bearing may be eliminated by suitably modifying the electrical properties of the receiver, and the latter is then said to operate as a sense-finder. Alternatively the ambiguity may be resolved by cross bearings from two or more stations.

The loop aerial

3. Consider a receiving aerial in the form of a rectangular conductive loop A B, C D, fig. 1, erected near to the surface of the earth in a vertical plane, its height being h metres and its width d metres. It is required to find the polar diagram of this loop for receiving purposes under the following conditions :—

- (i) The transmitter is situated at a distance which is very much greater than either h or d and also compared with the wavelength of operation.
- (ii) The wave front at the location of the receiver is a plane surface.
- (iii) The electrical field strength of the wave is uniform over the whole region embraced by the loop, whatever its orientation.
- (iv) The wave is normally polarized.

Under these conditions the electrical field vector at the receiver will be in the vertical plane and the magnetic field vector in the horizontal plane; these fields are of course, alternating quantities. The electric field strength will be denoted by $\gamma = \hat{r} \sin \omega t$ the unit being the volt

CHAPTER XVI.—PARA. 4

per metre. The vertical polarization of the electric field ensures that no E.M.F. will be produced in the horizontal sides of the loop, which serve merely to form a conductive path for any current which may flow. It is therefore necessary to consider only the E.M.F. produced in the vertical sides. Two special cases will first be taken.

Resultant loop E.M.F.

4. (i) When the plane of the loop is perpendicular to the direction of propagation of the wave, the vertical sides of the loop are equi-distant from the transmitter and the E.M.F. produced in each side will be $h\gamma$ or $h\hat{r} \sin \omega t$, that is, they will be in phase with each other. At a given instant, *e.g.* when the E.M.F. has its peak value $h\hat{r}$ in the upward direction in the wire A B, there will be an equal peak E.M.F. $h\hat{r}$, also in the upward direction, in the wire C D. These two E.M.F.s are therefore acting in opposition to each other and the E.M.F. available to drive current round the loop is zero.

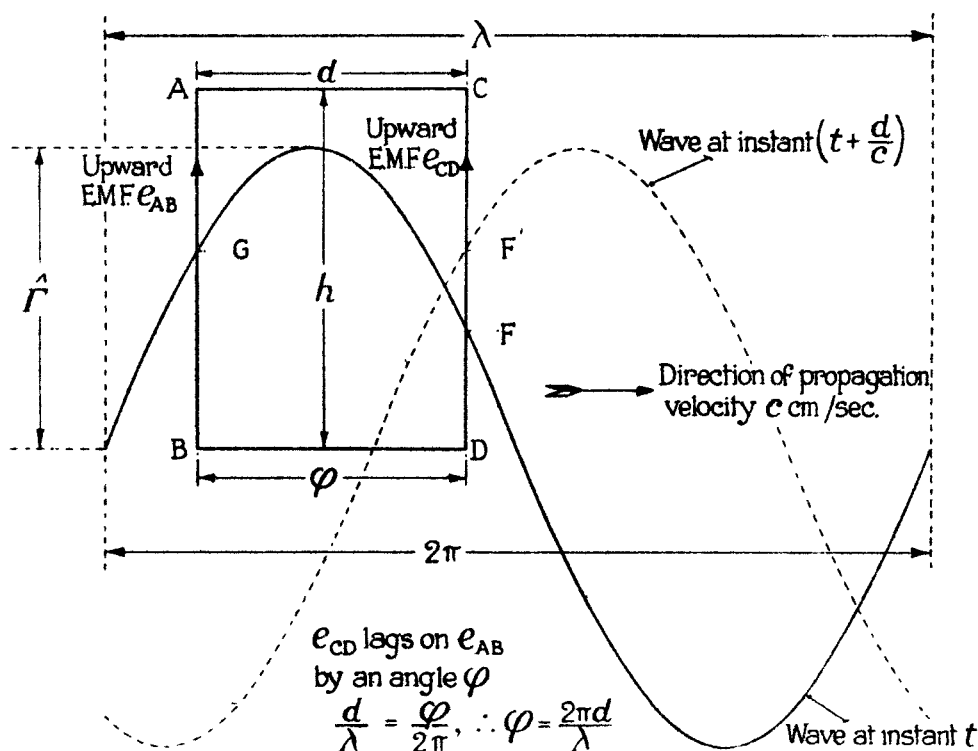


FIG. 1, CHAP. XVI.—Phase difference between E.M.F.s in opposite sides of loop aerial.

(ii) When the plane of the loop is parallel to the path of the wave, the peak value of the E.M.F. induced in each vertical side of the loop is again $h\hat{r}$, but the two E.M.F.s are no longer in phase, because at any given instant the two sides are not subject to the same portion of the electro-magnetic field. In the diagram, fig. 1, the strength of the electric field from point to point in space at a given instant, is shown by the solid-line sine curve. The instantaneous E.M.F. e_{AB} induced in the wire A B is proportional to G B, and that in the wire C D, *i.e.* e_{CD} , to F D. As the wave is travelling from left to right with a velocity $c = 3 \times 10^{10}$ centimetres per second, the field distribution a fraction of a second later will be as shown by the dotted-line sine curve. The E.M.F. e_{CD} induced in C D is now proportional to F' D, and

F'D = G B, *i.e.* e_{CD} at this instant is equal to e_{AB} at the previous instant. A moment's reflection will show that the instantaneous E.M.F. in C D will undergo exactly the same cycle of variation as that in A B, but these variations will be executed a fraction of a second later, *i.e.* the E.M.F. e_{CD} lags on the E.M.F. e_{AB} by some angle which may be denoted by φ . This angle is easily found from the diagram. If the wavelength λ and the width d of the loop are known, by simple proportion

$$\frac{\varphi}{2\pi} = \frac{d}{\lambda}$$

$$\text{and } \varphi = \frac{2\pi d}{\lambda} \text{ (radians).}$$

The total E.M.F. acting round the loop may now be deduced. The E.M.F.s in A B and C D are equal in peak value, but differ in phase by $\frac{2\pi d}{\lambda}$ radians. It is convenient to specify the phase of these E.M.F.s with reference to that which would be induced in a wire of equal length situated on the axis of the loop. If this E.M.F. is $e_0 = \mathcal{E} \sin \omega t$ where $\mathcal{E} = h\hat{r}$, the E.M.F. in A B will lead on e_0 by an angle $\frac{\pi d}{\lambda}$, while that in C D will lag by $\frac{\pi d}{\lambda}$. Hence

$$e_{AB} = \mathcal{E} \sin \left(\omega t + \frac{\pi d}{\lambda} \right)$$

$$e_{CD} = \mathcal{E} \sin \left(\omega t - \frac{\pi d}{\lambda} \right).$$

The total E.M.F. acting round the loop is $e_R = e_{AB} - e_{CD}$, or

$$e_R = \mathcal{E} \left\{ \sin \left(\omega t + \frac{\pi d}{\lambda} \right) - \sin \left(\omega t - \frac{\pi d}{\lambda} \right) \right\}.$$

By means of the trigonometrical identities

$$\sin (P + Q) = \sin P \cos Q + \cos P \sin Q$$

$$\sin (P - Q) = \sin P \cos Q - \cos P \sin Q$$

this expression simplifies to

$$e_R = 2\mathcal{E} \sin \frac{\pi d}{\lambda} \cos \omega t,$$

i.e. the peak value of the resultant E.M.F. is $2\mathcal{E} \sin \frac{\pi d}{\lambda}$.

Vector diagram

5. The above trigonometrical manipulation is illustrated by a vector diagram in fig. 2. The vector $\mathcal{E} = h\hat{r}$ represents the peak value of the E.M.F. which would be induced in a vertical wire of height h situated at the centre of the loop. Then $\mathcal{E}_{AB}$ is the corresponding value of the E.M.F. induced in A B, leading on the vector $\mathcal{E}$ by an angle $\frac{\varphi}{2} = \frac{\pi d}{\lambda}$ radians, while the corresponding E.M.F. in C D is represented by the vector $\mathcal{E}_{CD}$ which lags on $\mathcal{E}$ by $\frac{\varphi}{2}$ radians. The

CHAPTER XVI.—PARA. 6

resultant E.M.F. $\mathcal{E}_R$ is found by taking the vector difference between $\mathcal{E}_{AB}$ and $\mathcal{E}_{CD}$. It is seen that $\sin \frac{\varphi}{2} = \frac{\mathcal{E}_R}{2} \div \mathcal{E}_{AB} = \frac{\mathcal{E}_R}{2\mathcal{E}}$ because $\mathcal{E}_{AB}$ is numerically equal to $\mathcal{E}$ although they are not in phase, hence

$$\begin{aligned}\mathcal{E}_R &= 2\mathcal{E} \sin \frac{\varphi}{2} \\ &= 2h\hat{\Gamma} \sin \frac{\pi d}{\lambda}.\end{aligned}$$

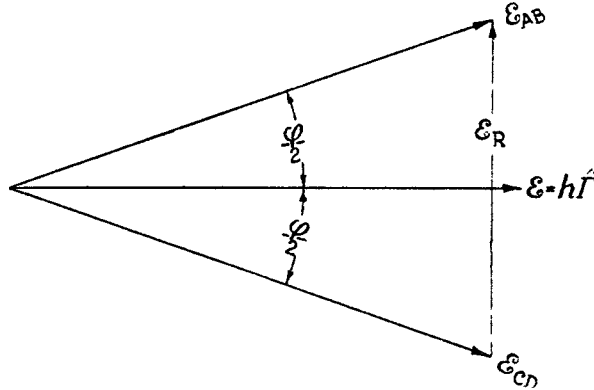


FIG. 2, CHAP. XVI.—Vector derivation of magnitude of resultant E.M.F. in loop aerial.

The instantaneous value of the resultant voltage is therefore

$$e_R = 2h\hat{\Gamma} \sin \frac{\pi d}{\lambda} \cos \omega t.$$

as already stated. The vectorial interpretation of the factor $\cos \omega t$ is that $\mathcal{E}_R$ leads on $\hat{\Gamma}$ by 90° .

Horizontal polar diagram

6. Having investigated these two special cases, let us now consider the general case, where the loop is neither in the plane of propagation nor perpendicular to it. Let the loop be so oriented as to make an angle θ with the direction of propagation, as indicated by the sectional plan, fig. 3. Then in passing from the vertical side A B to the vertical side C D, the wave must travel not d metres but $d \cos \theta$ metres. Hence the E.M.F. induced may be found by substituting $d \cos \theta$

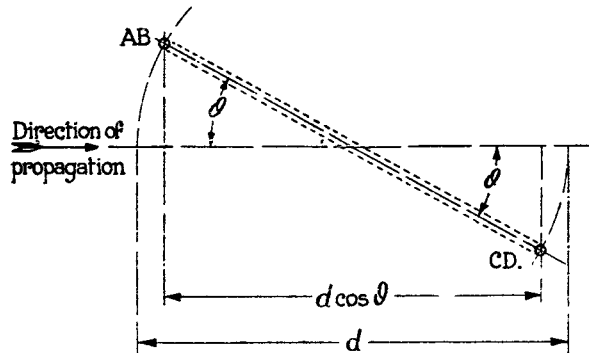


FIG. 3, CHAP. XVI.—Effective width of loop.

for d in the final equation of paragraph 5. If $e_{R\theta}$ is the resultant E.M.F. in the loop when it makes an angle θ with the direction of propagation

$$e_{R\theta} = 2h\hat{r} \sin \frac{\pi d \cos \theta}{\lambda} \cdot \cos \omega t.$$

Let us now examine the angle $\frac{\varphi}{2} = \frac{\pi d}{\lambda}$ for practical cases of loop reception. It is shown in

Chapter V that when $\frac{\varphi}{2}$ is a very small angle its magnitude (in radians) and its sine are practically equal. Taking a numerical example, for the reception of waves of the order of 500 kc/s (600 metres) we are not likely to use a loop wider than about 10 metres, in fact, it is usually convenient to use a considerably smaller one. In this case $\frac{\pi d}{\lambda} = \frac{3 \cdot 1416}{60} = \cdot 05236$ radians; the sine of $\cdot 05236$ radians differs from $\cdot 05236$ only in the fifth decimal place. For practical purposes, then, we may write

$$e_{R\theta} = 2h\hat{r} \times \frac{\pi d \cos \theta}{\lambda}$$

or, as $h\hat{d}$ is the area A of the loop

$$e_{R\theta} = \frac{2\pi A\hat{r} \cos \theta}{\lambda}.$$

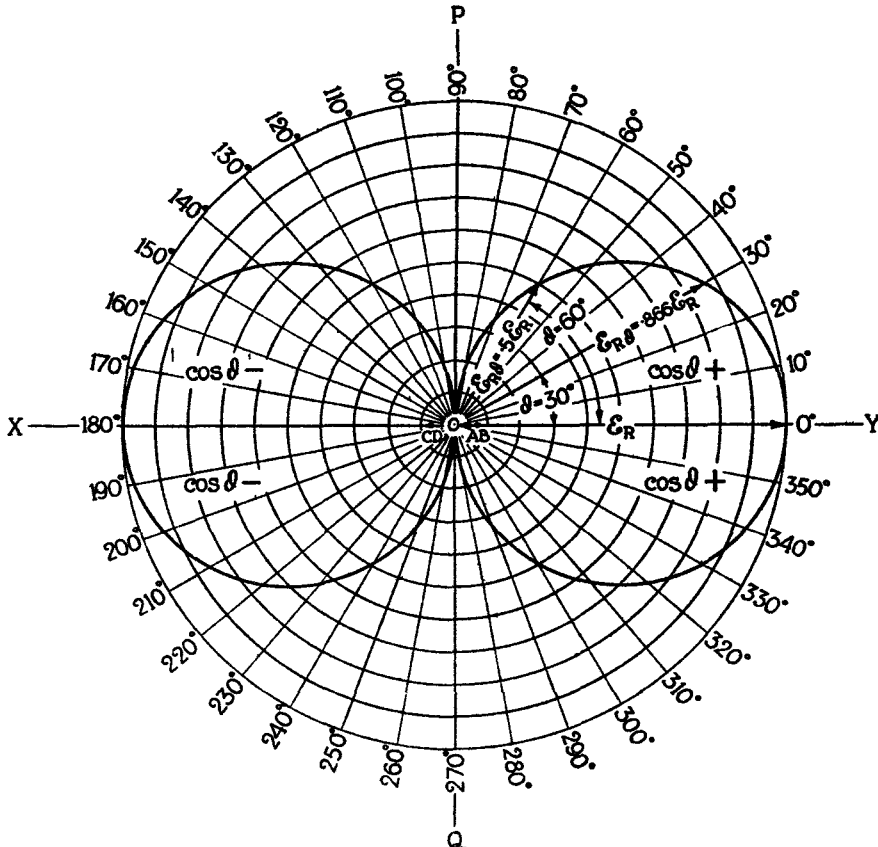


FIG. 4, CHAP. XVI.—Polar diagram of loop aerial (figure-of-eight diagram).

CHAPTER XVI.—PARA. 7

The horizontal polar diagram can now be plotted from this expression, and is shown in fig. 4, in which the vertical sides of the loop are denoted by A B and C D and the point O is the midpoint between them. Then O Y is the reference line from which angles must be measured, and may be drawn to represent $\mathcal{E}_R$ to any convenient scale. The angle θ is measured in an anti-clockwise direction from O Y. For example, if $\theta = 30^\circ$, $\mathcal{E}_{R\theta} = \mathcal{E}_R \cos 30^\circ$ or $0.866 \mathcal{E}_R$. Similarly if $\theta = 60^\circ$, $\mathcal{E}_{R\theta} = 0.5 \mathcal{E}_R$ and if $\theta = 90^\circ$, $\mathcal{E}_{R\theta} = 0$. In this quadrant, $\cos \theta$ is positive and the appropriate sign is inserted in the diagram. In the same way, the values of $\mathcal{E}_{R\theta}$ in the second and third quadrant may be drawn, but as $\cos \theta$ is negative between 90° and 270° , negative signs have been inserted in these quadrants, while between 270° and 360° $\cos \theta$ becomes positive once more. The polar diagram is seen to consist of two circles of diameter $\mathcal{E}_R$ and is often referred to as a figure-of-eight diagram. If the line P Q be supposed to represent the trace of a vertical plane perpendicular to X Y through the point O, the resultant E.M.F. set up by a transmitter situated to the left of P Q will act round the loop in a direction contrary to that set up by a transmitter situated on the opposite side, and this fact is of importance in the action of sense-finding devices.

Frame aerals

7. (i) The resultant E.M.F. acting in a single loop is very small compared with the E.M.F. in the vertical sides, but this may be partly overcome by using a number of turns of wire instead of a single loop. The resultant E.M.F. is then the sum of the E.M.F.s induced in the turns, and is therefore equal to the resultant E.M.F. of a single turn multiplied by the number of turns N , or

$$\mathcal{E}_R = \frac{2\pi N A \hat{f}}{\lambda}.$$

This form of aerial is usually referred to as a frame aerial. Two methods of construction are shown in figs. 5a and 5b. The first type, fig. 5a, is called a box frame, the turns being wound side by side, while the second, fig. 5b, is called a pancake frame, the turns being wound in a flat spiral. The pancake frame is often mounted, as shown, with one diagonal as the axis of rotation, but

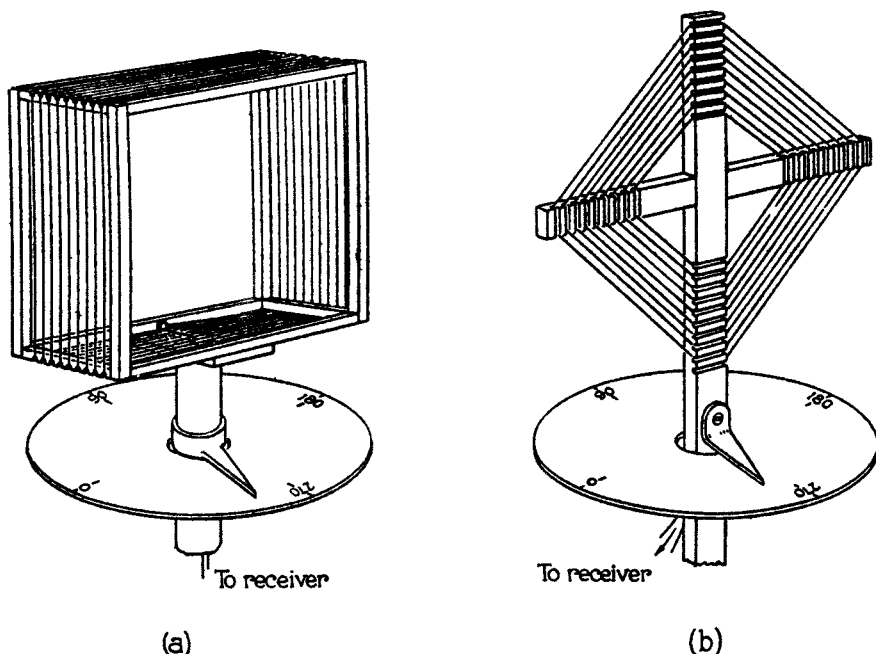


FIG. 5, CHAP. XVI.—Box and pancake types of frame aerial.

this does not reduce the resultant E.M.F., which can be shown to be proportional to the product of the area of the frame and the number of turns as in the former instance. In certain designs it is convenient to wind the frame in the form of a circle, because this shape lends itself to the design of a metal screen surrounding the coils. The purpose of such a screen will receive attention later.

(ii) When the operating frequency is below about 1,000 kc/s, the frame aerial is generally tuned by connecting a suitable condenser directly across the ends of the winding (*see fig. 7*). For higher frequencies, however, the aerial is usually coupled to the receiver by means of a special type of radio-frequency transformer. The secondary winding of the latter is tuned to the desired frequency by means of a variable condenser, the arrangement being analogous to the so-called aperiodic aerial in ordinary non-directional reception. The following table shows the order of the inductance and self-capacitance of different square, box frame aerials, each of which was designed to operate at frequencies below 1,500 kc/s.

h (metres).	No. of turns.	Spacing (cm).	L (μH)	C (μF)
1	8	·3	200	50
1·25	6	·6	150	60
2	4	·6	130	70
2·5	3	1·2	100	80

The given self-capacitance is only very approximate and is deduced from the fact that with no tuning capacitance the frequency of the frame aerial was found to be very nearly 1,600 kc/s in each case.

Effective height and pick-up factor of frame aerial

8. (i) The receptive property of a loop or frame aerial is often stated with reference to its effective height. This is the effective height of a vertical aerial in which a given field would induce an E.M.F. equal in magnitude to the resultant E.M.F. of the loop when oriented for maximum signal. Since the peak voltage induced in a vertical aerial of effective height h_e is $h_e \hat{E}$ volts, and the resultant E.M.F. in the frame aerial is $\frac{2\pi N A \hat{E}}{\lambda}$ volts, it follows that the effective height of the frame aerial is

$$h_e = \frac{2\pi N A}{\lambda} = \frac{\omega A N}{3 \times 10^8},$$

and is therefore dependent upon the frequency of the incoming wave. In all practical cases, the effective height is only a fraction of the actual height of the loop as defined in *fig. 1*. For example, in the case of a single turn 1 metre square, the effective height at a frequency of 300 kc/s is only 0·0628 metres. If the frame aerial is tuned to the desired frequency by means of a variable condenser connected across its ends, the peak voltage at the condenser terminals will be $\chi \mathcal{E}_a$, where χ denotes the circuit magnification as usual. The radiation resistance of a frame aerial is negligible compared with the radio-frequency ohmic resistance of the conductor, and there is no added resistance due to an earth connection. Thus the magnification of the frame aerial is generally very much greater than that of an open receiving aerial, and this compensates to some extent for the comparatively poor "pick-up" of the frame.

Example

A frame aerial 1 metre square having 20 turns, is tuned to 300 kc/s by a capacitance of 0·0003 μF . If its resistance at this frequency is 8 ohms, find the P.D. V_e set up at the condenser terminals by an R.M.S. field of 1 millivolt per metre, when the loop is in the "maximum" position.

CHAPTER XVI.—PARA. 9

The effective height of a single turn 1 metre square is $\cdot 00628$ metres, and of 20 turns is $\cdot 00628 \times 20 = \cdot 1257$ metres.

The resultant loop E.M.F. $E_s = h_e \Gamma = \cdot 1257$ millivolts (R.M.S.)

The circuit magnification $= \frac{1}{\omega C R} = \chi$.

$$\frac{1}{\omega C R} = \frac{1}{2\pi \times 300,000 \times \cdot 0003 \times 10^{-6} \times 8}$$

$$= 220 \text{ (approximately)}$$

$$V_c = \chi E_s$$

$$= 220 \times \cdot 1257 \text{ millivolts}$$

$$= 27\cdot 7 \text{ millivolts.}$$

(ii) In order to compare the receiving properties of different forms of direction finder, a quantity called the pick-up factor, p , is sometimes used. This is the ratio of the P.D., V , at the input terminals of the receiver proper, to the strength, Γ , of the electric field acting on the aerial system. Thus if a vertical aerial of effective height h_e and resistance R is tuned to a frequency $\frac{\omega}{2\pi}$ by the addition of an inductance L , and the P.D. across this coil is applied to the receiver,

$$V = \frac{\omega L}{R} h_e \Gamma$$

$$p = \frac{V}{\Gamma} = \frac{\omega L}{R} h_e.$$

Note that p is expressed in the same unit as h_e , i.e. in metres. The pick-up factor is in fact nothing more than the product of the circuit magnification and the effective height. The pick-up factor of a loop aerial tuned by a capacitance C across its terminals is equal to

$$\frac{2\pi AN}{\lambda} \times \frac{1}{\omega C R} \text{ or } \frac{AN}{3 \times 10^8 C R'}$$

C being expressed in farads, and R in ohms.

Maximum and minimum methods of obtaining bearing

9. A loop or frame aerial so mounted as to be capable of rotation about a vertical axis may be used to determine the position line upon which a distant transmitter is situated. As the loop is rotated, the signal strength will vary, reaching a maximum when the plane of the loop is coincident with the direction of propagation of the wave. Theoretically, the signal strength will fall to zero when the plane of the loop is perpendicular to the plane of propagation but for reasons discussed later, it is only rarely that an absolute zero is obtained. It would therefore appear that the bearing of the transmitter could be ascertained (with an ambiguity of 180°) by observation of the orientation giving either the maximum or minimum signal strength, as shown diagrammatically in fig. 6. The relative advantages and disadvantages of the two methods are as follows.

(i) Using the maximum method, if two operators are available, one may be engaged in finding the bearing while the other transcribes the W/T signal as received. This is an advantage in certain circumstances, e.g. in the case of an aeroplane transmitting an enemy report to a ground D/F station.

(ii) If the D/F receiver is located in a position where the interference level is high, it is difficult to observe the exact bearing upon which the minimum occurs.

(iii) With little or no interference the minimum is comparatively sharp, whereas the maximum is always very flat, i.e. the disadvantage of the minimum method, with a high interference level, applies to the maximum under all conditions.

On the whole, then, the minimum method is preferable, but in the presence of a high noise level the bearing is usually taken by swinging the loop through a small arc on either side of the minimum, observing the bearings upon which the signal is just audible. The bearing (or its reciprocal) is assumed to be the mean of these. The rotating loop is, as a rule, fitted with a

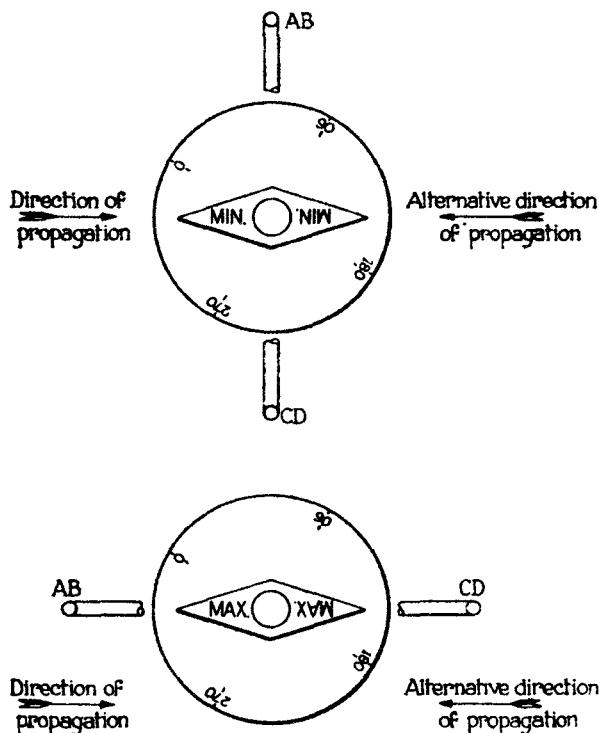


FIG. 6, CHAP. XVI.—“Maximum” and “minimum” methods of obtaining a bearing.

single pointer as in fig. 5, although it may appear that time would be saved by fitting a double pointer showing the bearing and its reciprocal at a glance. In operation however it is most desirable to turn the loop through more than 180° , observing both the minima. Owing to certain phenomena described below, these may not be exactly 180° apart.

Local errors

10. Before proceeding further with the practical application of the foregoing theory, it is necessary to deal with the phenomena which give rise to inaccurate bearings. Certain of these, which are due to the fact that the simplifying assumptions made in paragraph 3 are rarely applicable in their entirety, must be considered at a later stage, but immediate attention will be devoted to the errors which are due to the electrical properties of the direction-finding receiver itself. These errors may be attributed to

- (i) direct pick-up,
- (ii) vertical or antenna effect.
- (iii) displacement currents.

CHAPTER XVI.—PARA. 11

Direct pick-up in quadrature

11. The direct pick-up effect, as its name implies, is the result of the E.M.F. induced by the incoming electro-magnetic wave in any portion of the electrical wiring preceding the detector valve, or even in the post-detector stages if any electrical coupling exists between this portion of the receiver and the R.F. circuits. Referring to fig. 7, let A B, C D be a loop aerial with its plane in the direction of propagation of the signal, which is incident in the direction A C. The loop is tuned to the desired frequency by means of a variable condenser C_1 which is located inside the screening box containing the receiver proper. The loop is assumed to be connected to this condenser by a short length of twin flexible cable, which is screened from direct pick-up, but in order to reach the lower terminal of the condenser an additional conductor E F of length l metres is necessary; this is in series with A B, and for simplicity, suppose this wire to be placed vertically under the axis of rotation of the loop. If the dimensions of the loop are,

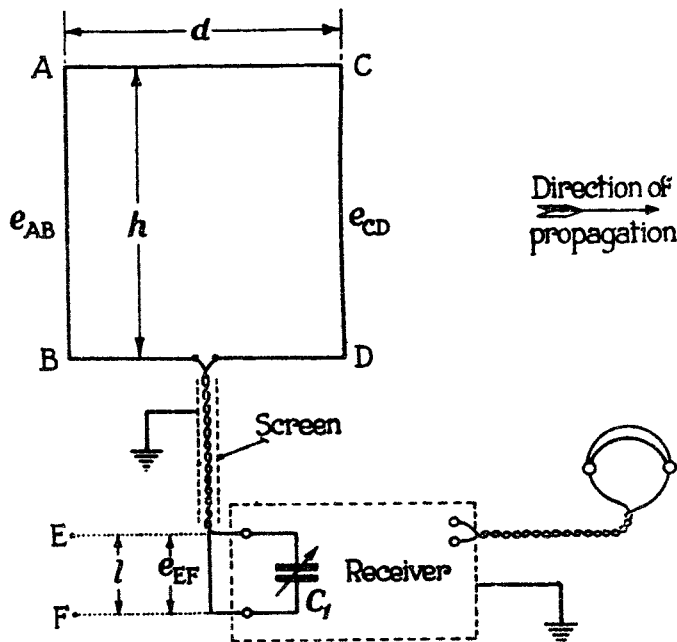


FIG. 7, CHAP. XVI.—Loop aerial with direct pick-up.

height h metres, width d metres as before, and the incoming wave has an electric field strength of peak value $\hat{F}$ volts per metre, the E.M.F. e_{AB} , set up in the side A B (fig. 7) will be

$$e_{AB} = h \hat{F} \sin \left(\omega t + \frac{\pi d}{\lambda} \right)$$

and in the side C D

$$e_{CD} = h \hat{F} \sin \left(\omega t - \frac{\pi d}{\lambda} \right).$$

As the conductor E F is also subject to the influence of the electric field, an E.M.F. e_{EF} will be set up in it. Since E F is in line with the axis of the loop, and the phase angles $\pm \frac{\pi d}{\lambda}$ are stated with reference to a conductor situated on this axis,

$$e_{EF} = l \hat{F} \sin \omega t$$

The resultant E.M.F. acting round the loop, in the absence of the conductor E F, has already been shown to be

$$e_R = 2h \hat{r} \sin \frac{\pi d}{\lambda} \cos \omega t.$$

Owing to the presence of the conductor E F, however, the total resultant E.M.F., e'_R , is $e_R + e_{EF}$ or

$$\begin{aligned} e'_R &= 2h \hat{r} \sin \frac{\pi d}{\lambda} \cos \omega t + l \hat{r} \sin \omega t \\ &= \frac{2\pi A \hat{r}}{\lambda} \cos \omega t + l \hat{r} \sin \omega t \\ &= \mathcal{E}_R \cos \omega t + \mathcal{E}_{EF} \sin \omega t. \end{aligned}$$

This resultant E.M.F. therefore consists of two components which are 90° out of phase with each other.

Effect of direct pick-up on polar diagram

12. The polar diagram of reception is plotted in fig. 8 for a particular ratio of $\frac{\mathcal{E}_R}{\mathcal{E}_{EF}}$. The figure-of-eight diagram shown in heavy line is the polar diagram of the loop alone, and the lighter circle is that of the wire E F, alone. Consider a wave incident in the direction P O. The peak

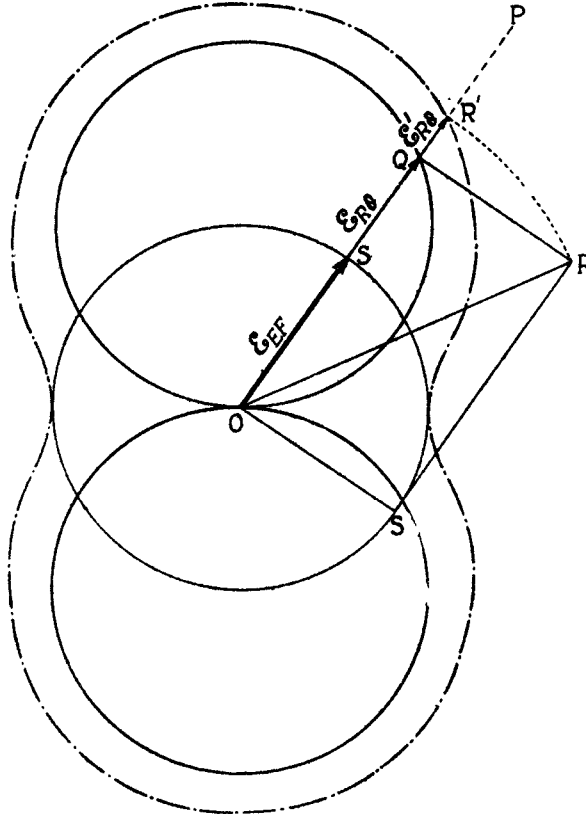


FIG. 8, CHAP. XVI.—Polar diagram showing effect of direct pick-up in quadrature with resultant loop E.M.F.

CHAPTER XVI.—PARAS. 13-14

value $\mathcal{E}_R$ of the resultant E.M.F. in the loop will be OQ , and the peak value of the direct pick-up E.M.F. $\mathcal{E}_{DP}$ will be OS ; since $\mathcal{E}_R$ and $\mathcal{E}_{DP}$ are in quadrature, the peak value of the resultant E.M.F. $\mathcal{E}'_{R\theta}$, due to a wave travelling along PO , will be $\sqrt{OQ^2 + OS^2}$. Graphically, $\mathcal{E}'_{R\theta}$ may be derived by drawing OS' , perpendicular to OS , and completing the parallelogram $OQRS'$. Then OR is the amplitude of $\mathcal{E}'_{R\theta}$, but its direction is along OQ and is easily transferred thereto, giving the point R' . The complete polar diagram is obtained by repeating this construction at 10° intervals and is shown in chain-dotted line. It is seen that the effects of any direct pick-up which sets up an E.M.F. exactly in quadrature with the loop E.M.F. are as follows:—

- (i) No orientation of the loop gives complete extinction of the signal; the strength of signals varies between a maximum and a minimum as the loop is rotated in azimuth.
- (ii) The minima are not sharp and are generally spoken of as "blurred minima."
- (iii) The minima are in their true positions with respect to the bearing of the transmitter, and are 180° apart, in accordance with the simple theory.

13. It is of interest to calculate the length l of unbalanced conductor which will cause the amplitude of the direct pick-up E.M.F. to be equal to that of the loop. Consider the loop previously taken as an example, (paragraph 8). It was there found that in a field having a strength of 1 millivolt per metre, the resultant loop E.M.F. was $\cdot 1257$ millivolt. The direct pick-up E.M.F. will be equal to this if $l = \cdot 1257$ metre or approximately 5 inches. Even such a short length of un-screened and unbalanced conductor will thus give an appreciable signal when the loop is in the

"minimum" position. The ratio of maximum to minimum signal is $\frac{\sqrt{(\mathcal{E}_R^2 + \mathcal{E}_{DP}^2)}}{\mathcal{E}_R}$. In the

above instance, therefore, the maximum signal is $\sqrt{2}$ times the minimum, corresponding to a change of only 3 db. In fig. 8, $\mathcal{E}_R = 2\mathcal{E}_{DP}$ and the ratio of maximum to minimum is $2\cdot 24$ to 1, corresponding to a change of about 7 db.

Direct pick-up in phase

14. The direct pick-up E.M.F. is rarely in exact quadrature with the loop E.M.F. as in the case just examined. For illustrative purposes, let us suppose the receiver in fig. 7 to be removed

in the direction AC to a distance of $\frac{\lambda}{4}$ from the centre of the loop, and connected as before by a

twin screened cable, with the exception of the unbalanced conductor EF , which is again vertical. The incident wave will now give rise to resultant E.M.F.s $\mathcal{E}_{DP}$ and $\mathcal{E}_R$ as before, but the field at EF will lag by 90° on the field at the axis of the loop and therefore $\mathcal{E}_{DP}$ and $\mathcal{E}_R$ will be either in phase or 180° out of phase with each other. If, in the vertical sides of the loop, the E.M.F. $\mathcal{E}_{AB}$ leads on $\mathcal{E}_{CD}$, i.e. if AB is the side nearest the transmitter, $\mathcal{E}_R$ and $\mathcal{E}_{DP}$ will be in anti-phase and vice versa. The total resultant E.M.F. $\mathcal{E}'_R$ is therefore equal to the sum of $\mathcal{E}_R$ and $\mathcal{E}_{DP}$ when the wave is incident in the direction CA , and to the difference between $\mathcal{E}_R$ and $\mathcal{E}_{DP}$ when the wave is incident in the direction AC . Referring to fig. 9, the polar diagram of the loop is shown by the figure-of-eight diagram, and that of the vertical aerial by a circle having the centre O . The polar diagram of the combination is found by algebraic addition of the two constituent diagrams, and is shown in chain-dotted line. If the wave is incident in the direction PO , the loop E.M.F. $\mathcal{E}_R$ is proportional to OQ , the direct pick-up E.M.F. $\mathcal{E}_{DP}$ to OR , and the total resultant E.M.F. $\mathcal{E}'_R$ to $OQ + OR = OS$. If incident in the direction $P'O$, however, $\mathcal{E}_R$ is proportional to $OQ' = -OQ$, $\mathcal{E}_{DP}$ to $OR' = OR$ and $\mathcal{E}'_R$ to $OQ' + OR = OR - OQ = -OS'$. The negative sign of the latter must be interpreted as signifying contrariety of direction with respect to the direct pick-up E.M.F.

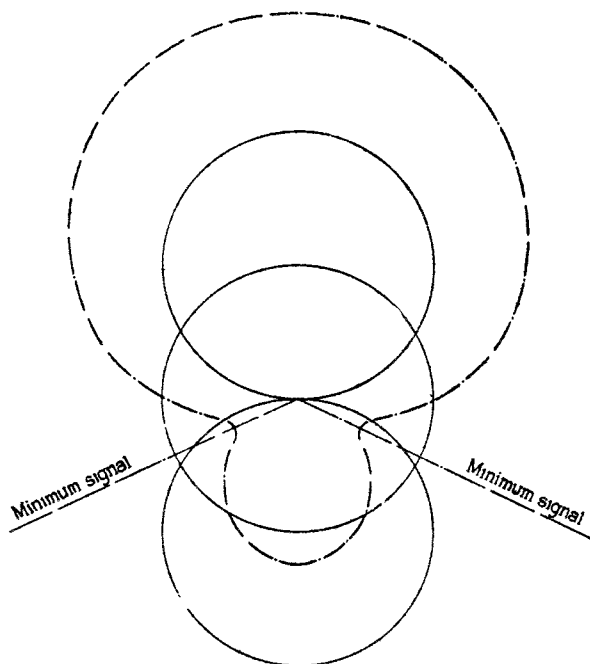


FIG. 10, CHAP. XVI.—Polar diagram showing general result of direct pick-up.

Elimination of direct pick-up

16. In the above discussion, the direct pick-up is assumed to occur in what is to all intents and purposes, a small vertical aerial directly coupled to the loop. Similar effects will however occur if the receiver is supplied with any E.M.F. due to the signal, other than the loop E.M.F. $\mathcal{E}_B$. To avoid any direct pick-up it is absolutely necessary to arrange all portions of the wiring symmetrically with respect to earth, and all connecting leads must be well screened by low-impedance conducting sheaths earthed at each end.

Vertical effect

17. "Vertical" or "antenna" effect is said to be present in a loop or frame aerial when unequal E.M.F.s exist in the vertical sides, due to imperfect electrical symmetry. The effect is therefore practically the same as direct pick-up, and both "in-phase" and "in quadrature" vertical effect may be found. As a rule, any accidental "vertical" is out of phase with the loop E.M.F. by an angle less than 90° . Vertical effect is often said to be due to the capacitance of the loop with respect to earth, but the significance of this statement is only appreciated after a careful examination of the conditions in which the resultant E.M.F. is obtained. In deriving an expression for the loop E.M.F. $\mathcal{E}_B$ in paragraph 4 it was assumed that the effective height h_e of each of the vertical sides A B, C D, of the loop is equal to the true height h . This is equivalent to assuming that the current is of the same amplitude at all points in the loop, and therefore that the wires have no capacitance with respect to each other or to the ground. This is of course incorrect, but provided that the two sides A B, C D, are perfectly symmetrical with respect to each other, and the ground, their effective heights will be the same and the induced E.M.F.s will have the same peak values, *i.e.* $\mathcal{E}_{AB} = \mathcal{E}_{CD}$. Let us now dispense with this assumption and

suppose the effective heights to be h_{AB} and h_{CD} respectively. If now a wave is incident along the plane of the loop in the direction A to C (cf. fig. 1) we have

$$e_{AB} = h_{AB} \hat{r} \sin \left(\omega t + \frac{\pi d}{\lambda} \right)$$

$$e_{CD} = h_{CD} \hat{r} \sin \left(\omega t - \frac{\pi d}{\lambda} \right),$$

or

$$e_{AB} = h_{AB} \hat{r} \left(\sin \omega t \cos \frac{\pi d}{\lambda} + \cos \omega t \sin \frac{\pi d}{\lambda} \right)$$

$$e_{CD} = h_{CD} \hat{r} \left(\sin \omega t \cos \frac{\pi d}{\lambda} - \cos \omega t \sin \frac{\pi d}{\lambda} \right).$$

Also

$$e_R = e_{AB} - e_{CD},$$

whence

$$e_R = (h_{AB} + h_{CD}) \hat{r} \sin \frac{\pi d}{\lambda} \cos \omega t + (h_{AB} - h_{CD}) \hat{r} \cos \frac{\pi d}{\lambda} \sin \omega t.$$

As already shown, the angle $\frac{\pi d}{\lambda}$ is always much smaller than unity. $\sin \frac{\pi d}{\lambda}$ may be replaced by $\frac{\pi d}{\lambda}$, and $\cos \frac{\pi d}{\lambda}$ by unity with negligible error, so that

$$e_R = \hat{r} \left\{ (h_{AB} + h_{CD}) \frac{\pi d}{\lambda} \cos \omega t + (h_{AB} - h_{CD}) \sin \omega t \right\}.$$

This E.M.F. is equal to that given by a loop having sides of equal effective height $h_0 = \frac{h_{AB} + h_{CD}}{2}$, with a length $(h_{AB} - h_{CD})$ of vertical unbalanced conductor situated on its axis. The above expression is therefore analogous to the final equation of paragraph 11. The polar diagram of a frame aerial with this form of vertical is similar to fig. 8 illustrating direct pick-up in quadrature.

Elimination of "vertical"

18. Having shown that "vertical" is due to lack of symmetry between the two sides of the loop, the steps necessary to reduce it to a minimum are easily seen. However carefully the loop itself and the connections thereto are arranged, one such asymmetry normally exists at the point

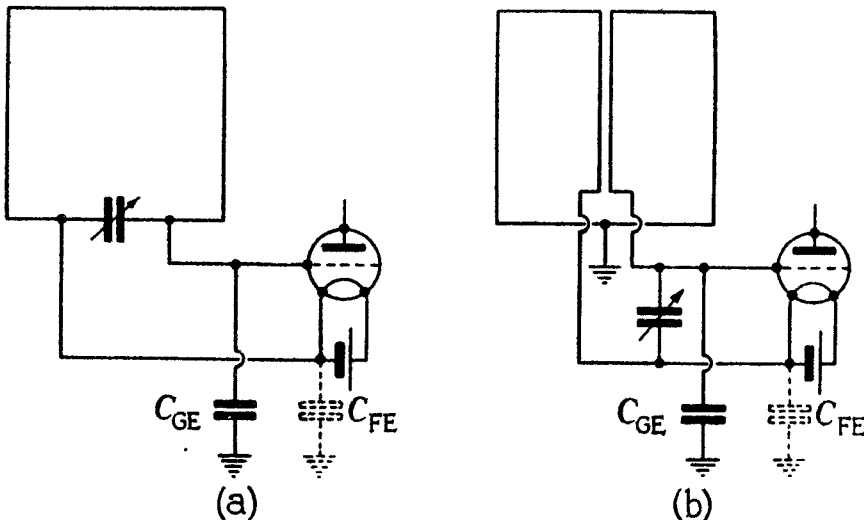


FIG. 11, CHAP. XVI.—Reduction of vertical effect due to unbalanced receiver.

CHAPTER XVI.—PARA. 19

where the resultant E.M.F.—or condenser P.D. in the case of a tuned loop—is applied to the first valve of the receiver. Referring to fig. 11a the filament of the valve is always at or near earth potential; even if insulated from earth, the capacitance C_{fb} between the L.T. battery and earth is invariably very much larger than the capacitance between the grid and earth. A possible remedy for this form of asymmetry is to connect a variable condenser C_{ab} between grid and earth, adjusting its value while taking a bearing, until the minima are 180° apart. In certain instances, and particularly on very high frequencies of the order of 10 Mc/s and above, it is necessary to achieve a high degree of symmetry and the early stages of the receiver may be arranged in push-pull to this end.

Screened loop

19. The effect of asymmetry caused by slightly unequal dimensions of nominally identical portions of a loop aerial can be reduced by winding the loop in two equal sections which are connected as in fig. 11b, the centre point of the loop being earthed. The capacitive balance just mentioned is still necessary for the purpose of balancing the amplifier input. The most effective method of reducing “vertical” in the frame or loop aerial itself is to ensure that the capacitance to earth of every element of wire in one side is exactly equal to that of the corresponding element in the other side of the loop. In this way the effective heights of both sides are equalized and vertical is practically eliminated. In practice this is often accomplished by enclosing each half of

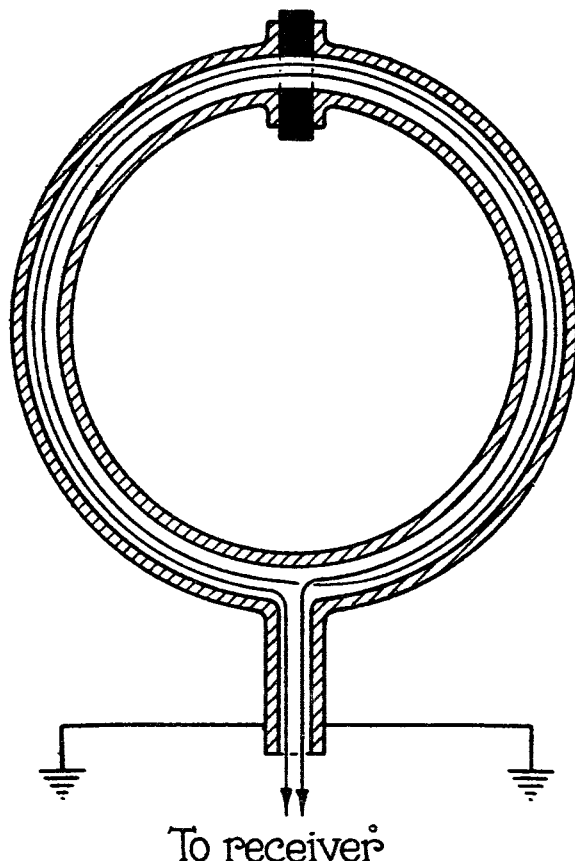


FIG. 12, CHAP. XVI.—Screened loop aerial.

the loop in a metal tube, the arrangement being referred to as a screened loop. This nomenclature is perhaps unfortunate since it is apt to give the impression that the loop is screened from the electro-magnetic wave, in which event no E.M.F. would be induced in it. Actually, the metal tube has negligible screening effect, because it does not form a closed path in which current can circulate. The most common method of construction is to use as a "screen" a metal tube bent into a circle, which is however broken at the centre point by the insertion of a tubular insulator of porcelain or of some phenolic compound (fig. 12), the tube itself being earthed at a point diametrically opposite to that at which the insulator is inserted.

Shielded R.F. transformer

20. In certain instances, the loop itself is aperiodic and is coupled to the receiver by means of a radio-frequency transformer in which the two windings are separated by an earthed electrostatic screen. The latter is so designed that although it is of metallic material, it does not form a closed circuit in which currents can circulate, and therefore does not screen the circuits from each other electro-magnetically. For low frequencies, e.g. below about 1,000 kc/s, a simple layer of copper foil may be employed, the overlapping edges being separated from each other by empire cloth. For higher frequencies a more elaborate construction is necessary. In one form, the primary winding is wound upon an ebonite former and a thin celluloid sleeve slipped over it. A close winding of silk-covered copper wire is put on this sleeve. All the turns are then bared for about one-eighth of an inch, a common earthing wire soldered to each, and the whole doped together with celluloid solution, the winding being then cut through at a point diametrically opposite to that at which the earthing wire is soldered. Finally another thin sheet of celluloid is doped down on the screen to serve as a foundation for the secondary winding.

Displacement effect

21. The error called displacement effect is always present in the box type of frame aerial. Suppose we have a frame consisting of four vertical conductors, A B, C D, A' B', C' D', suitably interconnected, as shown in perspective in fig. 13a. When the wave is incident in the direction shown, i.e. perpendicular to the plane of the loops, the E.M.F.s e_{AB} and e_{CD} neutralize each other, as do $e_{A'B'}$ and $e_{C'D'}$ likewise. The wires A B and A' B' also form a loop, however, the circuit being completed by the distributed capacitance between them. Similarly with the wires C D and C' D'. Referring to fig. 13b, it will be seen that these small "phantom" loops are so disposed that maximum resultant E.M.F. is induced when the plane of the frame aerial is perpendicular to the direction of propagation, i.e. when according to the foregoing theory, the resultant E.M.F. is zero. The displacement E.M.F. gives rise to errors resembling out-of-phase "vertical". Unlike the latter, however, the displacement E.M.F. changes sign when the frame is rotated through 180° .

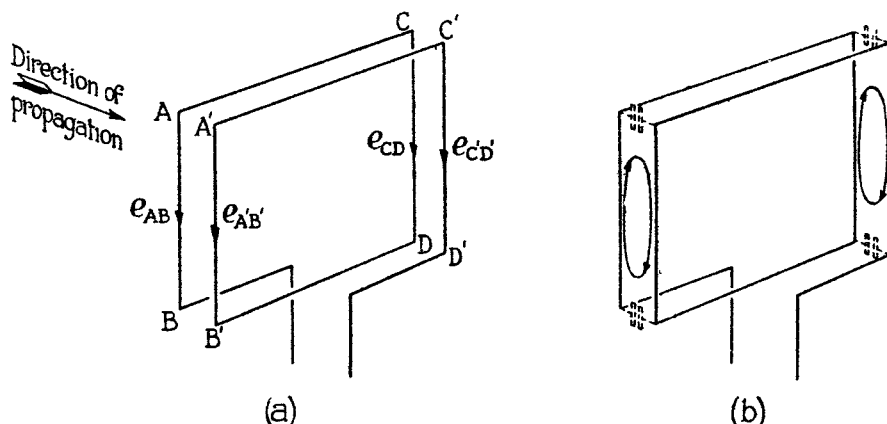


FIG. 13, CHAP. XVI.—Displacement effect in box-type frame aerial.

CHAPTER XVI.—PARAS. 22-24

Properties of box and pancake frames

22. (i) At first sight the absence of displacement effect would appear to make the pancake frame preferable to the box type. This is not necessarily so, because the former is inherently subject to appreciable vertical error. Referring to fig. 5 it will be seen that owing to the method of winding, the length of wire comprising one half-turn is necessarily longer than that of the corresponding half-turn on the other side of the frame. As a result, the effective heights of the two sides are not equal and the reasoning of paragraph 17 applies. This is what is meant by the statement that the effective capacitance of a pancake frame is greater than that of a box frame.

(ii) When it is necessary to use a frame consisting of a large number of turns it is usual to construct it by winding a number of pancake coils which are mounted side by side, and connected in series, thus forming a combination of the box and pancake types. The capacitance of such a frame is less than that of a pancake frame, while the product (number of turns $\times$ effective area) is greater. Since, for optimum resultant E.M.F. in the "maximum" position, this product must be as large as possible, the combined type gives a larger ratio of maximum to minimum signal than a pancake frame. The vertical and displacement errors are however less than those of a box frame having the same number of turns.

Coastal refraction

23. It is usually assumed that electro-magnetic waves arrive at the receiving station by the shortest path between the transmitter and receiver. This is not invariably the case. As stated in a previous chapter the velocity of the wave in or over a material medium is slightly less than in free, unbounded space. When a wave passes either over or through the boundary between two material media therefore, the direction of propagation may deviate through a small angle. One of the most important instances which arises in practical direction-finding is when the wave crosses the coast-line at an acute angle. As the velocity over sea is usually from 2 to 5 per cent. greater than over land, the direction of propagation may change by as much as 10° , while errors of from 3° to 4° are common. Where it is necessary to erect a ground D/F station near the coast, the arcs over which bearings are unreliable are usually noted during calibration and bearings lying in these arcs are treated with suspicion. The exact procedure to be adopted in such cases is a matter of signals organization. When possible however, such sites are avoided. Coastal refraction errors are of greater magnitude on high frequencies than on low and are generally very small on frequencies below 150 kc/s.

Sense-finding

24. The rotating loop or frame aerial gives a position line upon which the transmitter lies but it is often desirable to find the actual bearing of the transmitter, i.e. the direction in azimuth from which the wave reaches the receiving aerial. This is achieved by deliberately introducing a certain degree of in-phase vertical effect. Let us suppose that the relative magnitudes of the resultant loop E.M.F. and the vertical E.M.F. are such that the respective polar diagrams are those shown by heavy and light lines respectively in fig. 14, the relative direction of the E.M.F.'s being indicated by the conventional signs. Then the total polar diagram is found by adding the polar radii of the two diagrams, giving the curve shown by a chain-dotted line, which is called a cardioid or heart-shape. It will be seen that this polar diagram has only one minimum and one maximum. Both maximum and minimum are less sharp than are obtained with the figure-of-eight diagram, and the minimum is displaced by 90° from that of the figure-of-eight diagram. If then a rotatable receiving aerial system is so arranged that its polar diagram is heart-shaped, it can be used to determine the direction from which the wave is received i.e. it becomes a sense-finder.

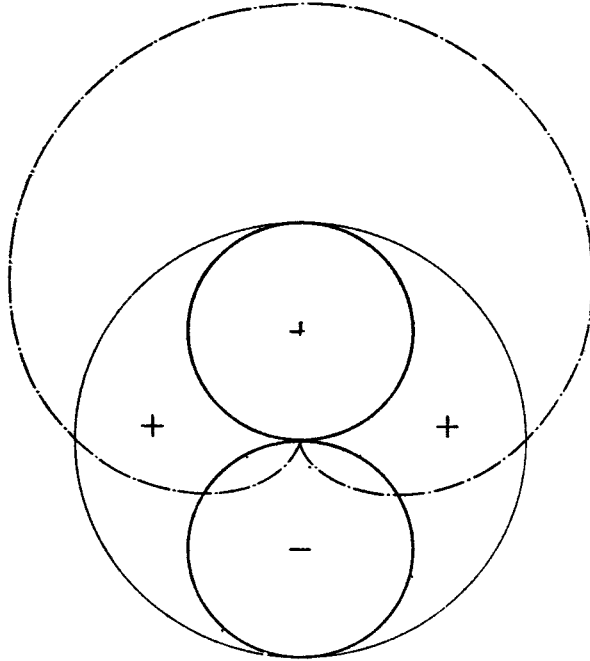


FIG. 14, CHAP. XVI.—Development of cardioid diagram.

Introduction of required "vertical"

25. The necessary vertical E.M.F. may be introduced in any one of several ways, provided always that the correct phasing is maintained. The arrangement shown in fig. 15a is most usually adopted on account of its simplicity. Here both the loop and the vertical aerial are inductively coupled to a tuned circuit, the P.D. across the condenser C being applied to the first valve of the receiver. A vector treatment will show whether the vertical and loop E.M.F.s cause in-phase P.D.s across the condenser. The relative phase of the condenser P.D. due to the

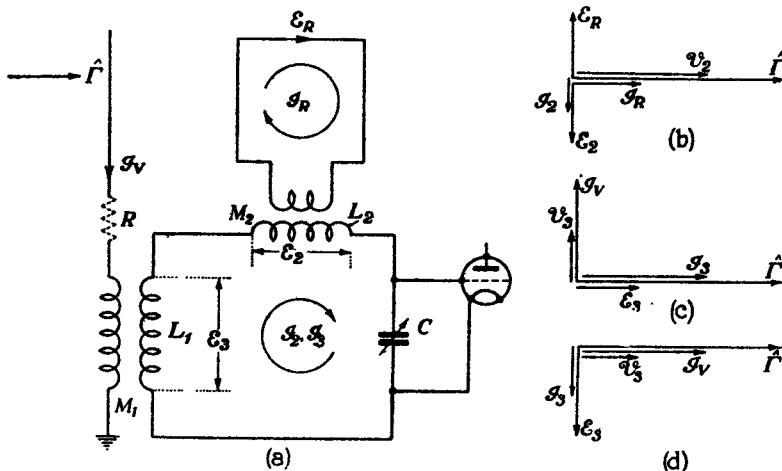


FIG. 15, CHAP. XVI.—Basic circuit for sense-finding and conditions for correct phasing.

CHAPTER XVI.—PARAS. 26-27

loop E.M.F. is shown in fig. 15b. The vector $\hat{I}$, which may be regarded as the datum, represents the electric field of the wave, and the resultant loop E.M.F. $\mathcal{E}_R$ is perpendicular to this (cf. paragraphs 4 and 5). Since the loop offers a positive (i.e. inductive) reactance, the loop current $\mathcal{I}_R$ will lag on $\mathcal{E}_R$ by 90° . The current $\mathcal{I}_R$ causes an induced E.M.F. $\mathcal{E}_2$, lagging on $\mathcal{I}_R$ by 90° , in the inductance L_2 . If the circuit L_1, L_2, C is tuned to the incoming frequency, the resulting current $\mathcal{I}_2$ is in phase with $\mathcal{E}_2$. The corresponding condenser P.D. $\mathcal{V}_2$ leads on $\mathcal{I}_2$ by 90° and is in phase with the original field.

Use of phasing resistance

26. (i) Now consider the vertical wire, the vector diagram being given in fig. 15c. The reactance of a short open aerial is predominantly capacitive, and the current $\mathcal{I}_V$ will lead on $\hat{I}$ by practically 90° . This current will induce in the inductance L_1 an E.M.F. $\mathcal{E}_3$, lagging on $\mathcal{I}_V$ by 90° , and a current $\mathcal{I}_3$ will flow in the tuned circuit, in phase with $\mathcal{E}_3$. The P.D. $\mathcal{V}_3$ set up between the condenser plates by this current will lead on $\mathcal{I}_3$ by 90° and will be in quadrature with $\mathcal{V}_2$. Hence this method of connection will not give in-phase E.M.F.s at the condenser terminals.

(ii) Now suppose that a resistance R is inserted in series with the vertical aerial, its value being very much larger than the capacitive reactance of the vertical wire. The conditions for the loop E.M.F. will not be affected, but the current $\mathcal{I}_V$ will now be practically in phase with $\hat{I}$, (fig. 15d). It follows that $\mathcal{V}_3$ will now be in phase with $\mathcal{V}_2$ and therefore the loop and vertical E.M.F.s combine in the correct manner to give a cardioid diagram of reception in azimuth.

(iii) Theoretically, the correct phasing could be achieved by tuning a very short vertical aerial to exact resonance with the incoming signal instead of by adding the resistance R . In these circumstances, however, a slight degree of mis-tuning is sufficient to swing the apparent sense through 90° in either direction, depending upon whether the aerial reactance becomes inductive or capacitive, and the method is therefore unreliable. With resistance phasing, if the circuit L_1, L_2, C is out of resonance, the voltages $\mathcal{V}_2$ and $\mathcal{V}_3$ are affected in the same manner and the sense-finding property is not impaired.

Value of phasing resistance

27. As a rule, the effective height of the vertical aerial is considerably greater than that of the frame aerial. To obtain a perfect cardioid diagram, the P.D.s $\mathcal{V}_2$ and $\mathcal{V}_3$ should be equal and the value of the resistance is chosen with this end in view. Referring again to fig. 15, if the inductance of the loop and the primary winding of the coupling M_2 is L , and the effective height of the vertical aerial is h_e , we have the following approximate relations:—

$$\begin{aligned}\mathcal{E}_R &= \frac{2\pi AN\hat{I}}{\lambda} = \frac{\omega NA\hat{I}}{3 \times 10^8} \\ \mathcal{I}_R &= \frac{\mathcal{E}_R}{\omega L}, \mathcal{E}_2 = \omega M_2 \mathcal{I}_R = \frac{M_2}{L} \times \frac{\omega NA\hat{I}}{3 \times 10^8} \\ \mathcal{E}_V &= h_e \hat{I} \\ \mathcal{I}_V &= \frac{\mathcal{E}_V}{R}, \mathcal{E}_3 = \omega M_1 \mathcal{I}_V = \frac{\omega M_1 h_e \hat{I}}{R},\end{aligned}$$

and a perfect cardioid will be obtained if $\mathcal{E}_2 = \mathcal{E}_3$,
that is if

$$\begin{aligned}\frac{\omega M_2 NA\hat{I}}{3 \times 10^8 L} &= \frac{\omega M_1 h_e \hat{I}}{R} \\ \text{or} \quad R &= \frac{3 \times 10^8 h_e L M_1}{N A M_2}.\end{aligned}$$

Thus, if the values of M_1, M_2, L and h_e are truly independent of frequency, the resistance required to give a perfect cardioid is constant, and can easily be found by calculation.

Example

In fig. 15, the vertical aerial has an effective height of 1 metre, the frame aerial has 20 turns of area 1 square metre, the total inductance being $200\mu H$, whilst $M_1 = M_2$. Find the resistance required to give a perfect cardioid.

$$R = \frac{3 \times 10^8 \times 1 \times 200 \times 10^{-6}}{20 \times 1} \\ = 3,000 \text{ ohms.}$$

Departure from perfect cardioid

28. In practice, the quantities M_1 , M_2 , L and h_e are not absolutely independent of frequency and it is not possible to obtain a perfect cardioid over a very wide frequency range with a single value of resistance. For a determination of sense, however, the perfect cardioid is not necessary, all that is required being an appreciable difference in the signal strength between the true bearing and its reciprocal. Fig. 16 shows the kind of polar diagram obtained when the vertical P.D. V_3 is considerably less than that required to give a true cardioid; the resemblance between this figure and fig. 9 should be noted. It will be observed that under these conditions, and in the absence of all other forms of pick-up, the polar diagram has two absolute zeros and two maxima,

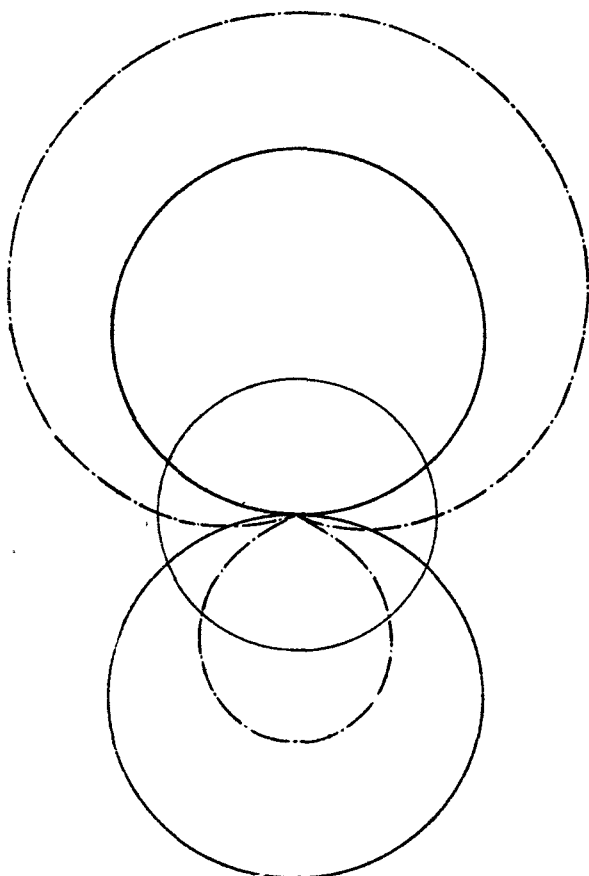


FIG. 16, CHAP. XVI.—Effect of reduction of vertical component.

CHAPTER XVI.—PARA. 29

one of which is greater than the other. A diagram of this kind is to be avoided if possible. If equality of the P.D.s $\mathcal{V}_2$ and $\mathcal{V}_3$ is not practicable it is preferable that the vertical P.D. should slightly exceed the loop P.D., giving rise to a diagram similar to that shown in chain-dotted line in fig. 17. This has only one maximum and one minimum and is less likely to lead to confusion. This requirement is sometimes met by providing a range of values of resistance.

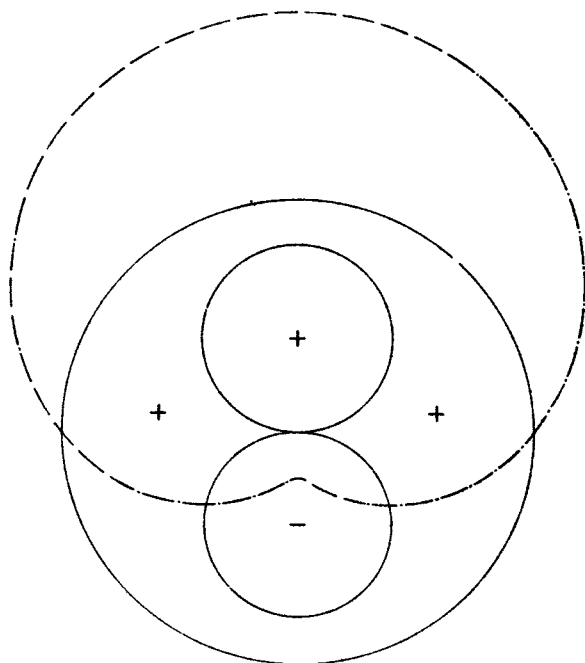


FIG. 17, CHAP. XVI.—Effect of increase of vertical component.

Adjustment of sense-finder

29. (i) In certain circumstances it may be possible to employ a continuously variable phasing resistance. The best adjustment for any particular frequency is then easily found in the following way. A double-pole two-way switch is fitted in the loop circuit, so that the connections of the loop to the coupling coil of the receiver may be reversed at will. With the vertical aerial disconnected and earthed, the loop is orientated to give maximum signal from a distant transmitter operating on the desired frequency. The vertical aerial is then connected in circuit, and the reversing switch is operated repeatedly while the phasing resistance is adjusted. The best value is obviously that which gives the greatest change of signal strength when the loop connections are reversed. In certain cases it is practicable and desirable to fit a separate pointer to the rotating loop in order to indicate the sense. Although for any given circuit it is possible to set this pointer from purely theoretical considerations, this procedure is seldom, if ever, adopted. The practice is to adjust it during the course of a test of the kind just described, i.e. by determining the sense of a transmitter in a known position. Observing however that if the D/F properties of the loop are accurate, the sense pointer must be at right angles to the pointer giving the D/F bearing it is possible to dispense with a sense pointer altogether. During the test the operator notes the actual bearing and the bearing shown by the D/F pointer; the former is either 90° more or 90° less than the latter, and may be recorded as "for sense add 90° to bearing shown" or "for sense subtract 90° from bearing shown". Provided that no connection is altered, this indication will then hold good on future occasions. It is therefore particularly necessary that the reversing

switch should be left in the position of calibration. Where a sense pointer is necessary, it is usually shorter than the D/F pointer and carries no definite indication of the exact bearing, in order to avoid any temptation to ascertain the latter by the sense-finding property alone. The bearing reported should always be one of the two obtained in the D/F position.

(ii) Although in the above discussion reference is made to the use of a vertical aerial, it must be understood that any form of aerial which has an approximately circular horizontal polar diagram, e.g. an L or T aerial, may be employed. Thus in an aeroplane, either the fixed aerial or the trailing aerial may be used for this purpose, the former being generally preferable.

AIRCRAFT D/F

Loop aerals

30. The installation and operation of D/F apparatus in aircraft presents a number of problems which are not met with in a ground D/F station. Many of these difficulties are of course common to all forms of aircraft W/T equipment, e.g. space and weight considerations and engine noise. The design of a rotating loop aerial for use in modern high-speed aeroplanes presents more difficulty than is apparent at first sight. A loop 18 inches in diameter, in a screening tube about $1\frac{1}{2}$ inches in diameter, may increase the drag by about 2 per cent, and will reduce the maximum speed in approximately the same degree. A reduction in the size of the loop will reduce the drag but will also reduce the pick-up factor of the loop. A possible solution is to use a comparatively

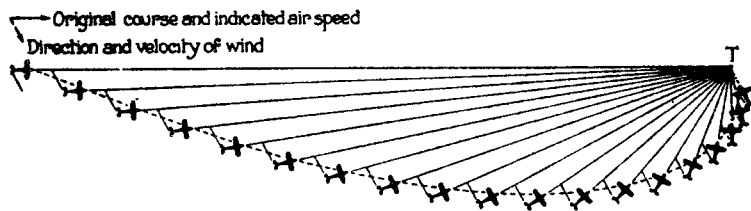
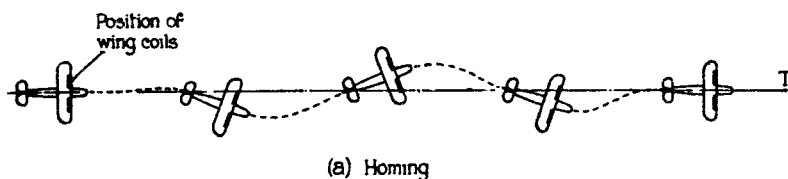


FIG. 18, CHAP. XVI.—Operation of wing coil D/F.

small loop, e.g. of about 12 inches diameter, mounted in a screened, streamlined casing above the structure. The simplest form of aeroplane D/F apparatus is that in which so-called wing coils are employed, but these are not suitable for use in all-metal aeroplanes. The wing coil is simply a frame aerial consisting of one or more turns of wire, the horizontal members being laid along and doped to the wing perpendicularly to the fore-and-aft line and the vertical members accommodated inside one of the struts on either side of the fuselage. Such a wing coil, when tuned by a variable condenser and connected to a suitable amplifier, will give minimum signals from a given transmitter when the aeroplane is heading directly towards or away from the transmitter. Since the 180° ambiguity is generally resolvable by geographic considerations, an installation of this kind may be used for homing. The method of operation is indicated in fig. 18a. When flying directly towards the transmitter station no signals are received, since the wing coil is perpendicular to the direction of propagation of the wave. The reception of signals is an indication that the

CHAPTER XVI.—PARA. 31

machine is off its course ; by swinging alternately to port and starboard, matching the strength of signals in each direction, the mid-point can be estimated and the course set as desired. If, however, a strong cross wind is blowing, homing by D/F bearings as above is subject to a certain limitation, owing to the drift to leeward of the machine. The effect of flying the D/F bearing without allowance for drift is shown in fig. 18b. In the original position, the course and speed of the aeroplane, and the direction and velocity of the wind, are as shown in the vector diagram. This course corresponds with the D/F bearing. After flying for a short distance on this course, the aeroplane will drift into the second position. If the pilot now corrects his course to the new D/F bearing, the aeroplane will, in a short time, reach the third position and so on. The actual course flown will therefore be as shown by the dotted line, and the aeroplane will reach the vicinity of the transmitter flying directly into the wind. The effect of drift must therefore be counteracted by setting a course slightly into the wind with reference to the D/F bearing after the latter has been ascertained.

Quadrantal correction

31. The wing coil system is not adapted for taking cross bearings in order to obtain a fix for which purpose a rotating loop is much more convenient. When used in an aeroplane, however, the rotating coil is subject to a type of error which is of no importance with the wing coil system. This error is caused by the induction of radio-frequency currents in the metal-work of the aeroplane itself. The existence of such currents connotes the production of both radiation and induction electro-magnetic fields, to which the D/F loop is subjected in addition to the radiation field of the distant transmitter. It must be realized that the fields due to the aeroplane itself are very complex owing to the different sizes, shapes and dispositions of the current-carrying members, but in general they may be resolved into two components which differ in direction in space and are in quadrature with respect to each other, as in fig. 19 in which the axis of the loop is located at the point P. The wave is incident in the direction T P, the positive direction of its electric field vector being upward out of the paper. For the purpose of illustration it is convenient to show the magnetic vector $\mathcal{H}$ which is positive in the direction shown, perpendicular to the direction of propagation. The instantaneous magnetic field of the incoming wave is of course sinusoidal and may be considered to be $\mathcal{H} \sin \omega t$ at the point P. The fields due to the aeroplane itself will then be (i) $\mathcal{H}_R \sin \omega t$, the radiation field, which is in phase with the wave field, and (ii) $\mathcal{H}_I \cos \omega t$, the induction field, which is in quadrature therewith. The direction in

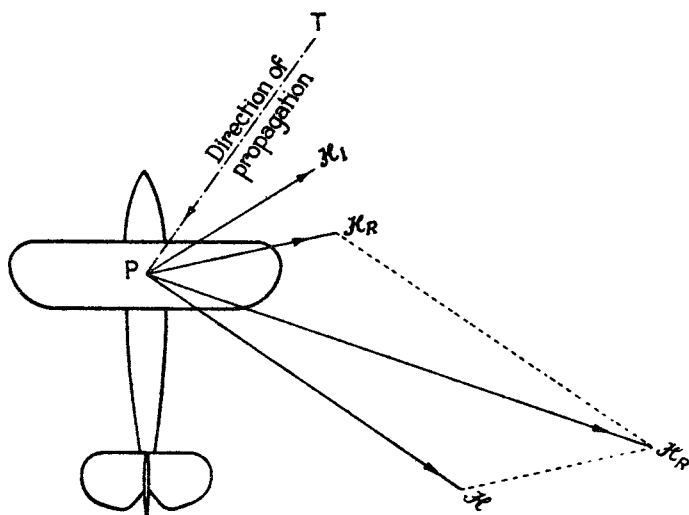


FIG. 19, CHAP. XVI.—Quadrantal error.

0	0	120	-9	240	+7
10	+6	130	-9	250	+5
20	+9	140	-10	260	+3
30	+10	150	-10	270	0
40	+10	160	-9	280	-3
50	+9	170	-6	290	-5
60	+7	180	0	300	-7
70	+5	190	+4	310	-9
80	+2	200	+8	320	-10
90	0	210	+10	330	-10
100	-3	220	+10	340	-9
110	-5	230	+9	350	-7

QUADRANTAL CORRECTION CARD

RAF. FORM 2026

QUADRANTAL CORRECTION CARD

AIRCRAFT TYPE:-

AIRCRAFT No:-

D.F. LOOP TYPE:-

D.F. LOOP SERIAL No:-

CHECKED:-

DATE:-

TRANSMITTER	FREQUENCY

INITIALS:-

RANK:-

FIG. 20

space of these fields depends upon the disposition of the current-carrying members of the aeroplane. Since the fields $\mathcal{H} \sin \omega t$ and $\mathcal{H}_1 \sin \omega t$ are in phase, they may be added, giving a resultant field $\mathcal{H}'_1 \sin \omega t$, so that finally we have two fields in quadrature, their magnitudes being $\mathcal{H}'_1$ and $\mathcal{H}_2$.

32. The effects of these fields may be considered separately. Supposing $\mathcal{H}_2$ to be zero, let the loop be rotated into the position giving minimum signal. This will obviously occur when the loop does not link with the field $\mathcal{H}'_1$, and the apparent bearing is then nearer to the fore-and-aft line than the correct bearing. If the aeroplane were perfectly symmetrical (a) on either side of the fore-and-aft line and (b) on either side of a transverse line through the point P, the error would be zero when the correct bearing was either 0° , 90° , 180° or 270° , and the maximum error would be found on bearings of 45° , 135° , 225° and 315° . On the other hand, if the field $\mathcal{H}'_1$ were entirely absent, maximum flux-linkage between the loop and the induction field $\mathcal{H}_2$ would occur when the plane of the loop was in the fore-and-aft line. Considering both fields together and dispensing with the assumption of a perfectly symmetrical aeroplane, it will be seen that as the fields $\mathcal{H}'_1$ and $\mathcal{H}_2$ are always in phase quadrature, and differ in magnitude and direction, they cannot be combined to give a resultant field having a constant direction in space. In general, their combination gives rise to a field which rotates in space at the frequency of the incoming wave and also varies in magnitude according to the bearing of the transmitter. As a result, the resultant field will affect the loop to some extent, in whatever direction it may be placed. The effect of the metalwork is therefore two-fold. In the first place it gives rise to erroneous bearings, maximum error usually occurring approximately in the middle of each quadrant. This is termed the quadrantal error. Its magnitude depends upon the construction of the aeroplane, and upon the position of the D/F loop; it is usually of the order of 6° to 12° . In the second place the presence of the induction field causes blurred minima, particularly when the bearing of the transmitter is in the neighbourhood of 90° or 270° .

Calibration

33. (i) The present practice is to allow for quadrantal error by calibrating the direction-finder. In brief, this process consists of taking D/F bearings of a transmitter of which the actual bearing is known, and so determining the correction to be applied. Since it is necessary to turn the aeroplane so that D/F bearings of the transmitter are obtained on a number of different relative bearings, the operation is sometimes termed "swinging for quadrantal correction" by analogy with the process of swinging an aeroplane in order to determine and correct the compass deviation. At the present time, however, no attempt is made to correct the loop by electrical means, although as mentioned later it is possible in certain cases to apply the correction mechanically. The term "relative bearing" used above and subsequently, denotes the bearing of a transmitter with reference to the head of the aeroplane, measured in degrees in a clockwise direction.

(ii) While it is possible to perform the operation of calibration in several ways, the following procedure is suggested as being as rapid as any, and requiring a minimum of calculation. It is applicable both to large and small aeroplanes. The preliminary step is to select a suitable transmitter, e.g. in the United Kingdom, one of the B.B.C. broadcasting stations, or any other which is known to maintain continuous transmission during the calibrating period. In certain circumstances, particularly overseas, it may be necessary to arrange for a suitable transmission in advance. The selected transmitter should, of course, give sufficient field strength at the aerodrome, but should not be too near the latter. Ideally, it should be from 50 to 100 miles away.

34. Calibration should be performed on a site as far away as possible from all buildings, the former practice of using the compass base for D/F calibration is only permissible if its site error (see paragraphs 40, 73, *et seq.*) is negligible. There is little to be gained by using it in any case. The D/F calibration when executed in the manner following depends upon either the pilot's or observer's compass in the aeroplane itself. It is, therefore, essential that the compass shall be

CHAPTER XVI — PARAS. 35-36

swung immediately prior to the D/F calibration. The occasions upon which the compass is swung are clearly defined in K.R. and A.C.I., and it is desirable to perform a D/F calibration as soon as possible after every occasion of such swinging. In any event it is important to ensure that all the removable equipment which is ordered to be in the aeroplane on the occasion of swinging for compass adjustment shall be in position during the D/F calibration. It must also be noted that the first fitting of a D/F loop may in certain circumstances entail compass swinging under these regulations.

35. In the actual calibration, it is preferable to refer only to magnetic and compass bearings, the magnetic bearing being the compass bearing corrected for compass deviation from Form 316 or Form 316a, depending upon which compass is employed. The *magnetic bearing* of the transmitter will be ascertained prior to the commencement of operations. If time permits, it is desirable to calibrate at intervals of approximately 10°, otherwise 15° or 20° intervals may be taken. The following pro forma should be prepared beforehand.

(1)	(2)	(3)	(4)
Line	Compass heading (proposed)	Compass heading (actually used)	Relative bearing of transmitter by D/F
1	0°		
2	10°		
3	20°		
4	30°		
5	40°		
6	50°		
7	60°		
8	70°		
etc.	etc.		

The figures in column 2 are intended merely as an aid to memory as to the proposed number of points, but it is not necessary to waste time in accurately placing the aeroplane on these bearings. For example, if the aeroplane is taxied on to the aerodrome, heading 42° (compass), column (3) of line 5 should be completed by inserting 42°, crossing out 40° in column (2). The relative bearing of the transmitter is ascertained and inserted in column (4). The aeroplane may then be so handled that the compass heading is 49°, columns (3) and (4) of line 6 being completed accordingly and the entry 50° in column (2) crossed out. Before taking the W/T observation on each heading the tail should be raised in order to bring the aeroplane into the normal flying position. In certain types of aeroplane, the error introduced by ignoring this instruction may be negligible but in the absence of definite orders to the contrary the tail should always be lifted.

36. When the data have been obtained for the required number of points, the following pro forma should be completed.

Magnetic bearing of transmitter320°.					
(1)	(2)	(3)	(4)	(5)	(6)
Line.	Head by Compass.	Magnetic Head	Relative bearing of transmitter	Relative bearing by D/F.	Quadrantal correction
1	0°	0°	320°	330°	-10
2	21°	22°	298°	306°	-8
3	39°	45°	275°	276°	-1
4	60°	64°	256°	252°	+4
5	81°	82°	238°	229°	+9
6	90°	90°	230°	220°	+10
etc.	etc.	etc.	etc.	etc.	etc.

Column (2) is identical with column (3) of the previous pro forma, while column (3) is column (2) corrected for compass deviation from Form 316 or 316a, depending upon which compass is used.

(1) Line	(2) Compass heading	(3) Magnetic heading	(4) Relative bearing of transmitter	(5) Relative bearing by D/F	(6) Quadrantal correction
1	0	0	320	330	-10
2	21	22	298	306	-8
3	39	45	275	276	-1
4	60	64	256	252	+4
5	81	82	238	229	+9
6	90	90	230	220	+10
7	100	98	222	212	+10
8	120	117	203	196	+7
9	140	136	184	183	+1
10	160	157	163	169	-6
11	180	180	140	160	-10
12	200	202	118	127	-9
13	220	224	96	98	-2
14	240	244	76	72	+4
15	260	262	58	49	+9
16	270	270	50	40	+10
17	280	279	41	31	+10
18	300	296	24	16	+8
19	320	315	5	3	+2
20	340	337	343	350	-7

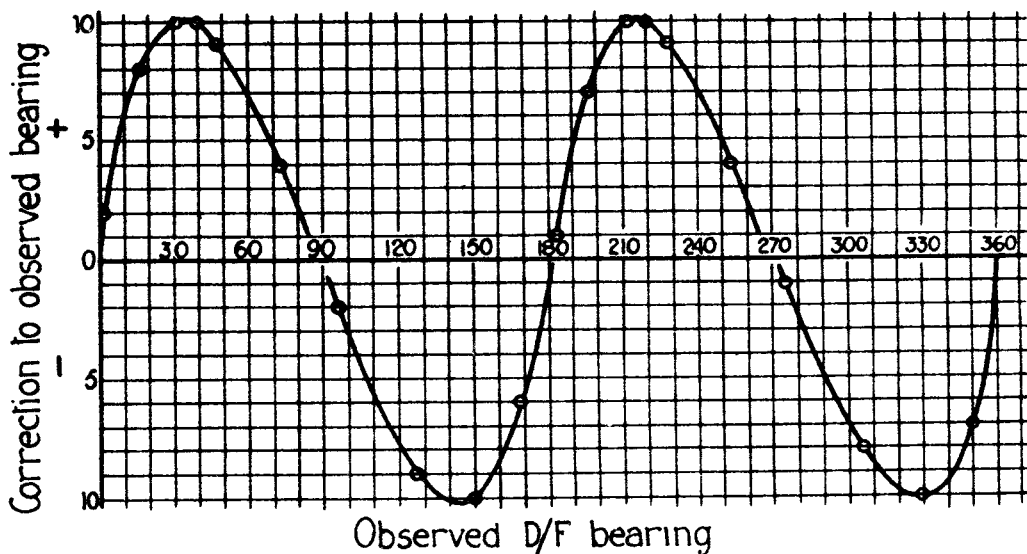


FIG. 21
PRO FORMA AND CORRECTION CURVE CHAP. XVI

Column (4) is obtained by subtracting the figure in column (3) from the magnetic bearing of the transmitter (adding 360° if necessary). Column (5) is identical with column (4) of previous pro forma. Column (6) gives the correction which must be applied to the W/T. bearing, Column (5), to obtain the relative bearing of the transmitter, column (4).

Quadrantal correction card

37. For purposes of easy reference, it is necessary to transfer the results of column (6) to the Quadrantal Correction Card, Form 2026. The latter is of the same shape and size as the Compass Deviation Card, Form 316a, and is intended to fit into a similar holder. An enlarged specimen copy is shown in fig. 20. It will be observed that the printed figures, 0 to 350 , show observed W/T bearings, and the quadrantal correction is to be inserted in the corresponding blank space. Since it is not expedient directly to obtain the relative bearings so printed, it is necessary to plot a curve showing the quadrantal correction for the whole 360° , with "D/F bearing" from Column 5 as abscissa and "quadrantal correction" from Column 6 as ordinate. A curve should be drawn through the points so obtained and the correction for 10° , 20° etc. taken from the curve and entered on Form 2026. Until further experience has been obtained, it is difficult to say to what extent it is permissible to smooth out this curve, but any point widely diverging from the general trend should be verified. In certain circumstances the graph may reveal the necessity for a constant, or nearly constant, correction on all bearings, in addition to the quadrantal correction. This correction may be applied by shifting the pointer with reference to the plane of the loop, if such provision is made. A typical completed pro forma, and the correction graph are shown in fig. 21; the corrections taken from the latter are those entered on the Quadrantal Correction Card in fig. 20.

Calibration on fixed base

38. In certain circumstances it may be convenient or desirable to calibrate on a fixed base similar to that used for swinging a compass. It is necessary to choose a suitable site, out of the path of aeroplanes taking off or landing, and remote from all buildings, railway lines etc., in accordance with paragraphs 73 *et. seq.* A peg is driven into the ground at the centre of the site and a circle of 35 or 40 feet radius marked out. The direction of true north is then obtained by means of a theodolite or by observation of the sun's passage across the meridian at local noon. The circle is divided into arcs of 20° , pegs are driven in and strings stretched across opposite ones. These should all intersect at the centre of the circle, corresponding radial lines are then marked out on the ground by digging narrow trenches which may be filled in with white-wash. The accuracy of the calibration will depend chiefly upon the care with which the base is marked out. To perform it, the aeroplane is placed over or parallel to each line in succession and a pro forma is filled in as before, except that all bearings are either relative or true. The calibration is therefore independent of the aircraft compass. The pro forma is as follows :—

True bearing of transmitter 330°				
(1) Line	(2) True heading	(3) Relative bearing of transmitter	(4) Relative bearing by D/F	(5) Quadrantal correction
1	0°	330	340	-10
2	20°	310	318	-8
3	40°	290	292	-2
4	60°	270	270	0
etc.	etc.	etc.	etc.	

Column (3) is the true bearing of the transmitter minus the true heading, column (2). The remainder of the work is as before.

CHAPTER XVI.—PARA. 39

Precautions

39. A quadrantal correction calibration made on the ground as above does not necessarily hold good when the aeroplane is in flight, but may be expected to do so under the following conditions :—

(i) The rotating loop is fitted *above* the structure of the aeroplane. If this is not so the ground calibration is not valid in the air and an air calibration becomes necessary. This requires steady flying and therefore good weather conditions, but otherwise presents little difficulty although of course some experience of calibration is required.

(ii) The site chosen for calibration is reasonably clear of metallic masses which have appreciable pick-up on the frequency involved. It is difficult to dogmatize in this respect, but the conditions stated in paragraphs 65 *et. seq.* are generally applicable.

(iii) The site error of the source used for calibration is known.

With reference to (ii) and (iii) a possible method of ascertaining whether any error is present is to make a preliminary investigation with a special portable direction-finder consisting of a small screened loop mounted on a wooden structure so that it has negligible quadrantal error.

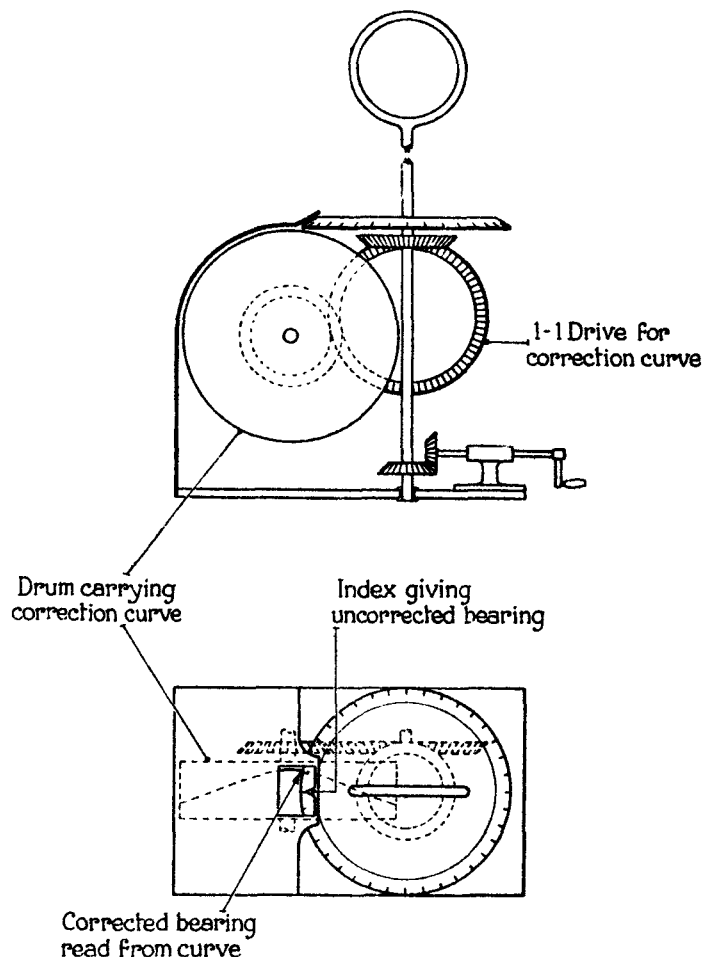


FIG. 22, CHAP. XVI.—Mechanical application of quadrantal correction.

Mechanical correctors

40. In certain installations, the quadrantal correction may be applied automatically. This may be achieved by incorporating a special cam in the drive between the shaft of the rotating loop and the bearing pointer. As initially supplied the cam is circular and the pointer gives the uncorrected reading. After calibration, the periphery of the cam is machined to shape in such a manner that on any given bearing, the pointer lags or leads on the loop to an extent governed by the required correction. Should this method be adopted, special instructions will be promulgated with regard to the operation of cutting the cam. In another form of automatic corrector, the bearing scale rotates with the shaft of the loop and the uncorrected bearing is read from a fixed lubber line engraved on a stationary bracket. A drum is geared to the shaft in such a manner that it makes one complete turn for each revolution of the loop. The axis of the drum is perpendicular to the shaft of the loop and its curved surface rotates in proximity to the periphery of the bearing scale. A line drawn on the curved surface may therefore be used as a lubber line for observation purposes, and this line may be so drawn as to show the corrected bearing. Both the methods described embody the same mechanical principles, and the latter is illustrated in fig. 23 since the mechanical details are less complicated than in the former type.

41. Where it is intended to operate the loop as a sense-finder, the occasion of calibrating for quadrantal error provides a suitable opportunity of adjusting the phasing resistance in order to obtain the best practicable cardioid, provided that the "vertical" is to be provided by a fixed aerial above the airframe. Obviously this is not possible if a trailing aerial is to be used for sense indication. It must always be remembered, also, that if provision is made for the use of either of two aerials, one above and one below the structure, the sign of the 90° correction to the D/F pointer (*see* paragraph 29) will change from positive to negative or vice versa on changing from one aerial to the other. Since their effective heights may be very different, a change-over from one to the other will also generally necessitate a re-adjustment of the phasing resistance.

Conversion angle

42. Radio waves travel over the earth's surface along arcs of great circles, but on charts and maps used for navigation, great circles are not represented as straight lines. Provided that both the transmitter and the aeroplane are in the area shown on a flying map, the difference between a straight line, and the arc of the great circle upon which they are situated, is negligible. Charts used for long distance flying are drawn on Mercator's projection, in which all meridians of longitude are parallel lines, instead of being convergent towards the geographical poles. On obtaining the bearing of a transmitter by means of a loop aerial, the navigator desires to draw, through the W/T station, a straight line upon which the aeroplane is situated, i.e. the position line referred to in paragraph 2. It must first be noted that the bearing from the W/T station is the reciprocal of the bearing of the latter from the aeroplane, but before drawing the position line upon the chart a correction must be applied to allow for the difference between great circle and Mercatorial bearings. This correction, which is called the conversion angle, is negligible in latitudes below 60° unless the distance between transmitter and aeroplane is more than 100 miles, but should always be applied in other cases. In order to preserve a record of the observation and to facilitate the work of obtaining the mercatorial bearing of the transmitter from the observed D/F bearing, Form 2058 has been introduced. This form will be issued in pads for insertion in a suitable cover, the inner side of which is finished in a washable white material. This is intended for recording in pencil the names, frequencies, and geographical positions of such W/T stations as are likely to be available for D/F purposes during a particular flight. It is suitably ruled and engraved with the appropriate legends for this purpose.

43. A specimen copy of Form 2058 is shown in fig. 23. The W/T operator is only concerned with the insertion of the name of the transmitting station, its frequency, and the following details of the bearing, viz. (i) the observed bearing (ii) the time at which it was taken (iii) the classification, as laid down in the signals organization. The time of observation is of primary importance

CHAPTER XVI.—PARA. 44

and must never be omitted. The remainder of the form (including the portion for quadrantal correction) is for the use of the navigator, but will be briefly summarized by an example, in order that W/T personnel may appreciate the significance of their work.

Example

During a flight from Reykjavik to Plymouth, a D/F bearing of 335° is obtained at 1035 hours from Aberdeen W/T station transmitting on 315 kc/s, when the true course is 140° and the position by dead reckoning is Lat. $60^\circ 32' \text{ N.}$, Long. $15^\circ 40' \text{ W.}$ The position of the W/T station, namely, $57^\circ 8' \text{ N.}$, Long. $2^\circ 4' \text{ W.}$, has previously been inserted in the cover for Form 2058.

The operator, after inserting the name and frequency of the transmitting station, enters the observed bearing, the time at which it was taken, and the classification of the bearing. He then passes the form to the navigator, who completes the work as follows:—

Time 1035 hours	Observed bearing	335°
G.M.T.	Quadrantal correction	$- 8^\circ$
	Corrected bearing	327°

The operator then passes the form to the navigator, who completes the work as follows:—

Latitude of W/T station	$57^\circ 8' \text{ N.}$
D.R. latitude at 1035	$60^\circ 32' \text{ N.}$
Sum latitudes	$117^\circ 40' (\doteq 118^\circ)$
Longitude of W/T station	$2^\circ 4' \text{ W.}$
D.R. longitude	$15^\circ 40' \text{ W.}$
D. Long.	$13^\circ 36'$
True Course	140°
True bearing of W/T station	467°

The latter is the sum of the true course and the corrected relative bearing; when this sum exceeds 360° , the latter figure is subtracted, giving

True bearing	107°
----------------------	-------------

To obtain the mercatorial bearing, the conversion angle must be applied. The necessary calculations have been made and the results exhibited in the form of a nomogram which is printed on the left-hand side of Form 2058, together with instructions for its use. A straight line drawn on the nomogram, through the points "D. Long., $13^\circ 36'$ " and "sum latitudes, 118° " gives the conversion angle, from the middle scale, as $\pm 6^\circ$ approximately. Since the aeroplane is west of the transmitter the negative sign is taken. Hence the final entry is

Mercatorial bearing	101°
-----------------------------	-------------

The reciprocal of 101° being 281° , the latter is the bearing on which the position line must be drawn from the transmitter.

The Robinson system

44. (i) Owing to the high noise level which prevails in an aeroplane, it is more difficult to obtain a reliable bearing by the minimum method than in a ground D/F station. In the Robinson system, an attempt is made to secure the advantages of both maximum and minimum methods. In this system two frame aerials are employed; these are rigidly fixed at right angles to each other in such a manner that the combination can be rotated in azimuth. The frame aerials are connected to the receiver proper as shown in fig. 24. The smaller aerial L_1 is called the main coil and the larger one, L_2 , the auxiliary coil. The relative proportions of these coils will be discussed later. Two change-over switches, S_1 , S_2 , are included. When thrown over to the left, the switch S_1 connects the auxiliary coil in series with the main coil, but when thrown over to the right, the auxiliary coil is removed from the circuit and a small coil L_3 is substituted. This coil is designed to have negligible pick-up, so that when it is in circuit, the signal is received only on the main coil. The inductance of L_3 is equal to that of the auxiliary coil. The aerial circuit is tuned to the desired frequency by a variable condenser C . This adjustment holds good for

	TRANSMITTING STATION <i>Aberdeen</i>		FREQUENCY 315 KC/S
	NOTE:- DO NOT CALCULATE OR APPLY THE CONVERSION ANGLE AT LATITUDES BELOW 60° UNLESS THE RANGE EXCEEDS 100 MILES		LATITUDE OF W/T STATION $57^{\circ} 8' N$ D.R. LATITUDE + $60^{\circ} 32' N$ SUM LATITUDES $117^{\circ} 40'$
	TO FIND C.A. :- PLACE RULER SO AS TO JOIN D.LONG. TO SUM LATITUDE. C.A. IS SHOWN WHERE RULER CUTS MIDDLE SCALE		LONGITUDE OF W/T STATION $2^{\circ} 4' W$ D. R. LONGITUDE ~ $15^{\circ} 40' W$ D. LONG. $13^{\circ} 36'$
	TIME 1035 HRS. G.M.T. <i>BBT. LOCAL</i>	OBSERVED BEARING 335 QUADRANTAL CORRECTION $\pm$ 8 CORRECTED BEARING 327	
	CLASS OF BEARING :- <i>First</i>	TRUE COURSE + 140 TRUE BEARING OF W/T STATION 467 IF NECESSARY, SUBTRACT 360° 360 TRUE BEARING 107	
	AIRCRAFT WEST OF TRANSMITTER SUBTRACT <i>ADD</i>	IF NECESSARY, APPLY C.A. $\pm$ 6 MER. BEARING 101	

STATION	FREQUENCY	LATITUDE	LONGITUDE	REMARKS
<i>Aberdeen</i>	315 KC/S	$57^{\circ} 8' N$	$2^{\circ} 4' W$	

Inside of cover

PRO FORMA FOR D/F CALCULATIONS

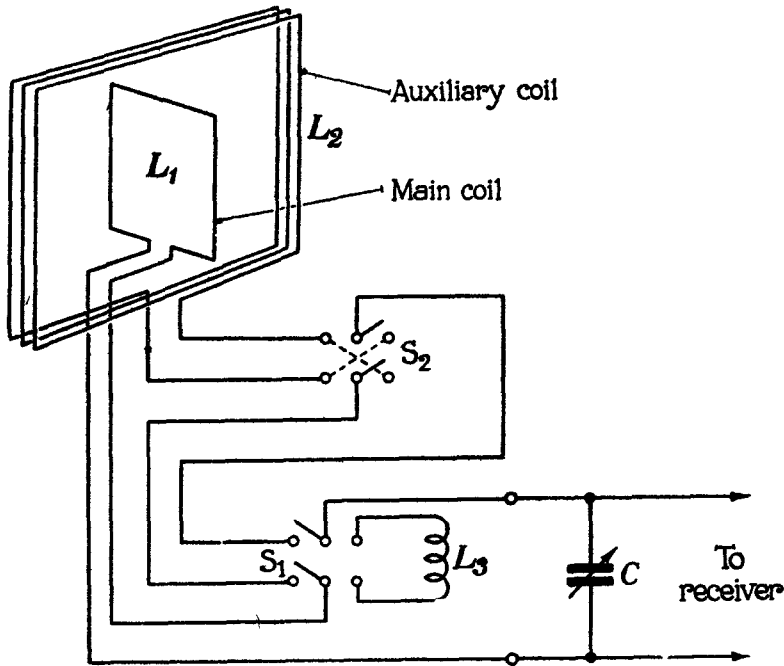


FIG. 24, CHAP. XVI.—Principle of Robinson system.

either position of the switch S_1 , provided that the inductance of L_3 is correctly adjusted. The switch S_2 is called the reversing switch. Its function is to reverse the direction of winding of the auxiliary with respect to that of the main coil, when both are in circuit.

(ii) The method of obtaining a bearing is as follows. The switch S_1 is thrown over to the right so that only the main coil is operative, and the coils rotated until the signal appears to be of maximum strength. Owing to the wide arc of maximum signal this bearing may be several degrees in error. If this is so, an E.M.F. is also set up in the auxiliary coil, although as this is not in circuit it has no effect upon the signal strength. If the switch S_1 is now thrown over to the left, both the main and auxiliary coils are in circuit. The resultant E.M.F. of the auxiliary coil is now algebraically added to that of the main coil and the signal strength will be either increased or decreased, according to whether the respective voltages are in phase or in opposition. It follows therefore that the signal strength will vary between two levels when the switch S_2 is thrown from left to right and vice versa. When the main coil is truly aligned in the direction of the transmitter, however, no resultant E.M.F. is developed in the auxiliary coil and the operation of the change-over switch S_2 has no effect whatever upon the signal strength. The operator therefore swings the coils through a small arc while reversing the switch S_2 as necessary until a position is found in which the signal strength is unaffected by the reversal. The bearing of the transmitter, with an ambiguity of 180° , is then given by the orientation of the main coil.

45. In order to obtain a high degree of accuracy, it is necessary that when the auxiliary coil is only very slightly off the zero-signal bearing, the resultant E.M.F. induced in it shall be sufficient to cause a perceptible difference in signal strength when its reversal is performed. It is difficult to say what constitutes a perceptible change, but for the present purpose we may

CHAPTER XVI.—PARA. 46

assume that 3 db. will be sufficient. Let us suppose the main coil to be θ degrees off the correct bearing. The resultant E.M.F.s in the two coils will then be

$$e_a = (A_a N_a K \sin \theta) \cos \omega t$$

$$e_m = (A_m N_m K \cos \theta) \cos \omega t$$

where e_a and e_m are the resultant E.M.F.s, A_a , A_m the areas, N_a , N_m the numbers of turns in the auxiliary and main coils respectively, and K is a constant. The amplitude of the total resultant E.M.F. is therefore

$$e_1 = (A_m N_m \cos \theta + A_a N_a \sin \theta) K$$

$$\text{or } e_2 = (A_m N_m \cos \theta - A_a N_a \sin \theta) K$$

depending upon the position of the reversing switch. The output of the receiver due to the above voltage will depend upon the nature of the rectification process. For C.W. reception with a strong heterodyne, (which is the normal method in modern practice in order to secure the maximum signalling range), the rectification is approximately linear. The change of signal strength

on reversal will then be proportional to the ratio $\frac{e_1}{e_2}$, that is to

$$\frac{A_m N_m \cos \theta + A_a N_a \sin \theta}{A_m N_m \cos \theta - A_a N_a \sin \theta}$$

For a change of 3 db. $\frac{e_1}{e_2} = \sqrt{2}$, and therefore

$$A_m N_m \cos \theta + A_a N_a \sin \theta = \sqrt{2} (A_m N_m \cos \theta - A_a N_a \sin \theta)$$

$$(1 + \sqrt{2}) A_a N_a \sin \theta = (\sqrt{2} - 1) A_m N_m \cos \theta$$

$$\frac{A_a N_a}{A_m N_m} = \frac{.414}{2.414} \cot \theta$$

$$= .172 \cot \theta.$$

From the final equation it is easily found that to obtain the required change of signal strength with a bearing error of 1° , the area-turns of the auxiliary coil must be 9.8 times that of the main coil, for 2° , 4.9, and so on.

46. Let us consider the conditions which prevailed at the time this system was in general use, namely the reception of spark or I.C.W. signals with little or no regenerative amplification. The detection process then approximates to square-law rectification and a change of 3 db. would

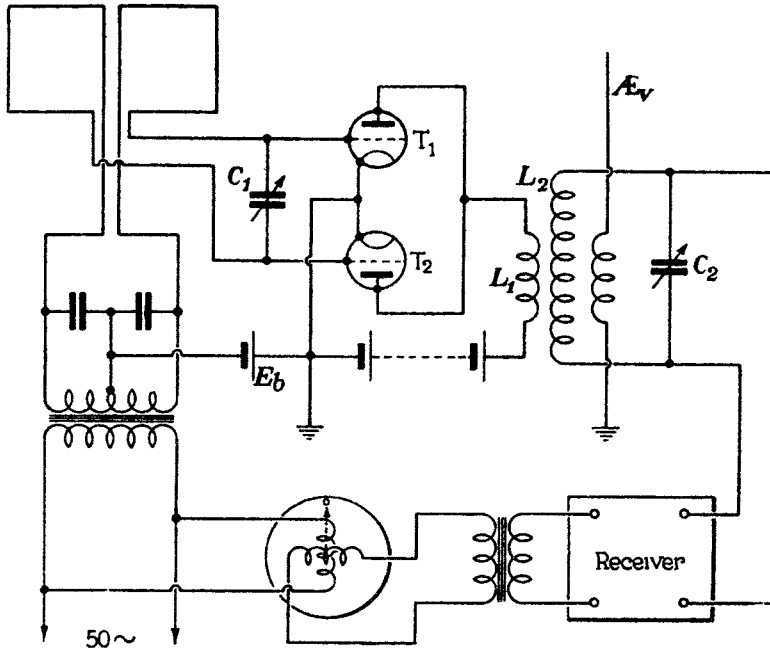
be obtained if the ratio $\left(\frac{e_1}{e_2}\right)^2$ were equal to $\sqrt{2}$, i.e. if $\frac{e_1}{e_2} = \sqrt[4]{2} = 1.19$.

Repeating the above calculation it then appears that $\frac{A_a N_a}{A_m N_m} = \frac{.19}{2.1} \cot \theta = .087 \cot \theta$. For

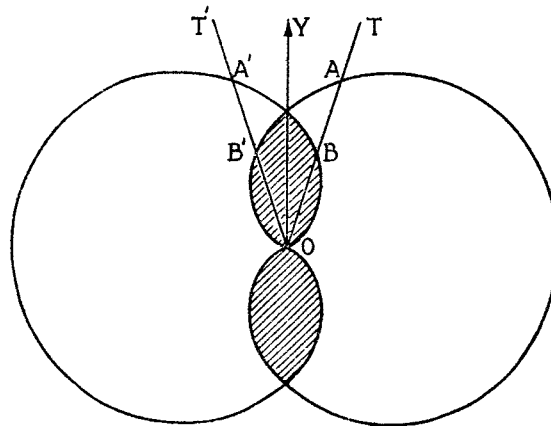
a discrimination within 1° , the ratio of auxiliary to main area-turns is then only 5, while a ratio of 2.5 gives a discrimination within 2° . In practice, ratios of from 2.5 to 6 were used for reception under the latter conditions. The system has fallen out of use owing to the large ratio required under modern conditions of reception. As the ratio is increased the system approaches more nearly to a simple frame aerial operated upon the minimum method, the so-called "main coil" merely contributing a small "displacement" component. Another disadvantage attending an increase in the size of the auxiliary coil is the difficulty of preserving a suitable ratio of inductance to minimum capacitance, in order to cover the desired frequency range. Finally, the adoption of all-metal construction precludes the use of wing coils, for which the system was originally designed.

The radio compass

47. The radio compass is intended chiefly for use in homing, and is a modification of the Robinson system. A simple form is illustrated in fig. 25a. The loop aerial is usually fitted in a suitable streamlined housing, in such a manner as to give minimum signal when the machine is flying towards the transmitter. The loop is wound in two equal sections, the electrical centre being connected to the common filament connections of the valves T_1 , T_2 , to the grids of which the outer ends of the two loops are also connected; a mean grid bias is established by the grid



(a)



(b)

FIG. 25, CHAP. XVI.—Principle of radio compass (fixed loop type).

CHAPTER XVI.—PARA. 48

battery E_b , this voltage being just sufficient to reduce both anode currents to zero. In addition to the steady bias, a 50-cycle alternating voltage is applied to the grids so that they are made alternately more or less negative, and each valve will pass anode current only under the latter condition. As the anodes of the two valves are in parallel, only the one having the less negative bias is operative at any given instant, and this valve supplies a radio frequency current to the coil L_1 ; consequently this current suffers an abrupt reversal in phase one hundred times per second, i.e. at each half-cycle of the alternating bias frequency, and the induced E.M.F. in the tuned secondary circuit $L_2 C_2$ will suffer a like reversal. This circuit is also coupled to a vertical aerial $\mathcal{A}$, which supplies an E.M.F. to the circuit $L_2 C_2$, in phase or in anti-phase with that due to the loop aerial, according to which of the two valves is passing anode current. It will be observed that the algebraic sum of the loop and vertical E.M.F.s will give rise to a cardioid diagram, (fig. 25b). The rectified current from the W/T receiver is supplied to the moving coil of a small dynamometer instrument, the fixed coil of which is synchronously energized at the bias frequency. Referring to fig. 25b, suppose the transmitter to be in the direction T, then the current in the moving coil of the dynamometer will be proportional to $OA - OB$, and the pointer will be deflected accordingly, while if the transmitter is in the direction T' the current will be proportional to $OB' - OA'$, and will be in the reverse direction. If however the transmitter is situated at Y, the current in the moving coil will be zero and the pointer undeflected. When flying approximately toward the transmitter, the application of right rudder causes the pointer to deflect to the right and vice versa, while if flying away from the transmitter, right rudder gives left deflection of the pointer. In practice therefore the 180° ambiguity is easily resolved.

48. A radio compass operating upon an entirely different principle is shown diagrammatically in fig. 26. This type consists of a small screened loop which is mounted in a suitable streamlined housing above the structure of the aeroplane, and is rotated at a constant speed of 300 r.p.m. by a small electric motor. The loop is connected by slip rings to the input circuit of an amplifier. If

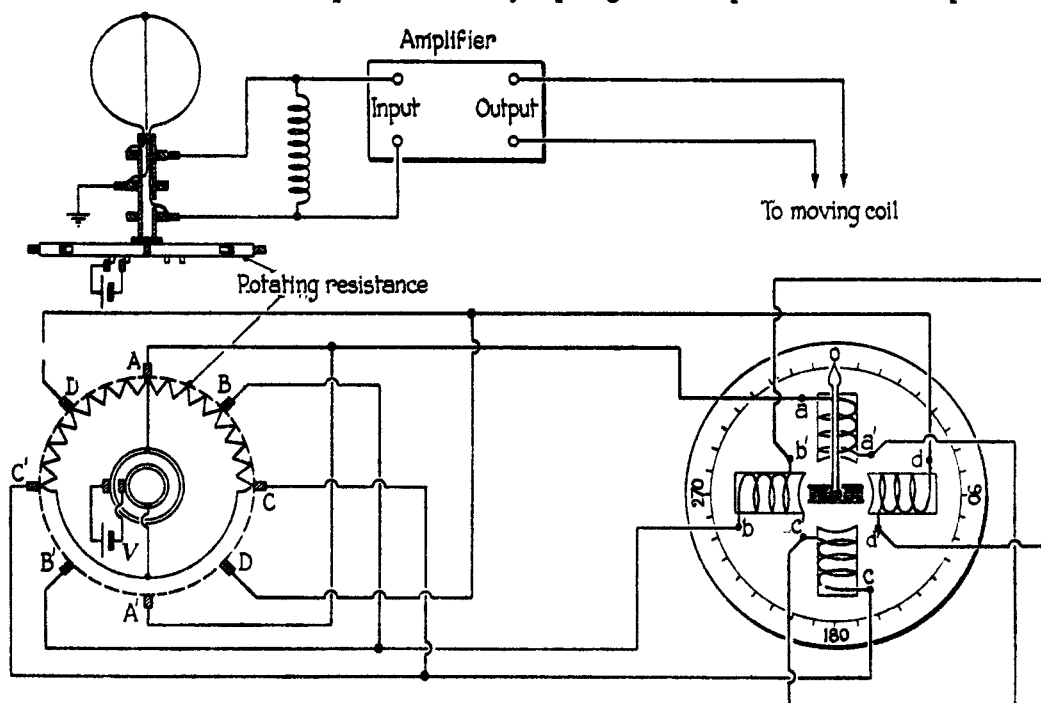


FIG. 26, CHAP. XVI.—Radio compass (revolving loop type).

a continuous wave of constant amplitude is received, the input to, and therefore the output of, the amplifier will vary according to the relative position of the loop with reference to the direction of the transmitter. To all intents and purposes, the output of the amplifier is an alternating current having a frequency of 10 cycles per second. This output current is caused to give an indication of the bearing of the transmitter in the following manner. The rotating spindle of the loop also carries on its lower extremity a semi-circular resistance, wound upon a circular former which rotates with the loop. This resistance is supplied with direct current from a source of constant E.M.F., and eight brushes are arranged at equal intervals round the former, opposite brushes being directly inter-connected, as shown in the diagram. The pairs A A', C C' form the terminals of one pair of supply leads and the pairs B B', D D' the terminals of another pair of supply leads for the bearing indicator. The latter is a particular form of dynamometer instrument.

49. Referring to the diagram it is seen that when the resistance is in the position shown, the brush A is V volts positive with respect to the brushes C, C', and a current will flow in the fixed coils a a' c c' of the indicator from a to c'. The brushes B and D' are at the same potential and no current flows in the fixed coils b b' d d' of the indicator. One-eighth of a revolution later, however, brush B is at a potential of V volts above that of the brushes D D', while the brushes A and C are at equal potential. After one-quarter of a revolution, the brush C is at a potential of V volts above that of the brushes A A', while the brushes B and D are at equal potential. Tracing a complete revolution of the resistance in this way it will be seen that the arrangement is in effect a two-phase generator. The fixed coils a a' c c' and b b' d d' of the indicator each carry an alternating current at a frequency of 10 cycles per second (two cycles per revolution), but the respective currents are in quadrature with each other. These currents therefore set up a rotating field in the air gap in which the moving coil of the dynamometer is situated. This field rotates in space 10 times per second.

50. The moving coil carries the output current of the amplifier, and during the reception of a wave of constant amplitude, from a fixed direction with reference to the aeroplane, the current in the moving coil will reach a maximum when the loop is pointing toward or away from the transmitter. At this moment, then, the total flux in the air gap will have its greatest possible value. The coil tends to turn to the position in which it is threaded by the greatest possible flux, and therefore takes up a position depending upon the direction of the distant transmitter. The coil is fitted with a pointer which indicates the direction of the transmitter upon a fixed scale graduated in degrees with reference to the fore-and-aft line, subject to an 180° ambiguity. The bearing indicator is lined up with the two-phase current generator by shifting the latter upon its spindle until correct bearings are obtained from a known transmitter. It is, however, necessary to apply a quadrantal correction, after calibrating, in the same manner as with the simple rotating loop. This correction is usually applied by means of a cam incorporated in the driving mechanism of the loop.

51. In the present stage of development all visual-indicating radio compasses suffer from the great disadvantage that they cannot completely discriminate between the wanted signal and that from a station on the same or an adjacent frequency, but on a different bearing. When such interference is present the instrument may either integrate the total field and give an erroneous bearing, or may oscillate violently and indicate no bearing whatever. It must be appreciated that such interference may be present even if it gives little or no audible indication on an aural receiver, e.g. it may be within the limits of the dead space, or sufficiently different in frequency to give a heterodyne beat above the audible limit. These instruments function best on a wave of constant amplitude such as the unmodulated carrier of an R/T transmitter, but are very erratic when the transmitter is deeply modulated or key-controlled. The indications are also unreliable unless the aeroplane is in perfectly steady flight.

GROUND STATION D/F

Bellini-Tosi system

52. It has already been shown that in all practical cases, i.e. when the width d of a loop aerial is small compared with the wave-length, the resultant loop E.M.F. is very much smaller than that set up in a vertical wire having a height equal to the actual height h of the loop. In order to obtain appreciable pick-up, we may increase the area of the loop, or the number of turns, or both, but practical limitations arise as follows. (i) An increase either in area or in the number of turns increases the inductance of the loop and therefore affects the frequency coverage. (ii) An increase in the number of turns increases the vertical and displacement effects, according to the manner of winding. (iii) An increase in the area necessitates a corresponding increase in the inertia of the frame, and it becomes too unwieldy for the rapid swinging necessary to obtain a bearing by the minimum method. Again, if an attempt is made to increase the effective operational range of a small frame aerial by additional radio-frequency amplification, a limit is imposed by the increased noise level. The Bellini-Tosi system was originally developed in order to overcome the above limitations.

53. The Bellini-Tosi direction-finder employs a system of fixed loop aerials in conjunction with a device known as a radio-goniometer. The principle of this instrument is shown in fig. 27; it consists of two small frame coils mounted rigidly at right angles to each other. These are called the fixed coils. A smaller coil, called the search coil, is mounted upon a spindle passing through the axis of symmetry of the fixed coils. The ends of this winding are connected to the receiver proper, while the ends of the fixed coils are connected to the loop aerials as described below. The whole assembly is completely screened by enclosure in a copper-lined box or compartment. A bearing plate is provided, and the spindle of the search coil carries a pointer which shows the angle at which the search coil is set with reference to one of the two fixed coils. The term "goniometer" means "angle-measurer." As the fixed coils are perpendicular to each other, the mutual inductance between them is zero, while the mutual inductance between either of the fixed coils and the search coil is variable, according to the angle at which the latter is set.

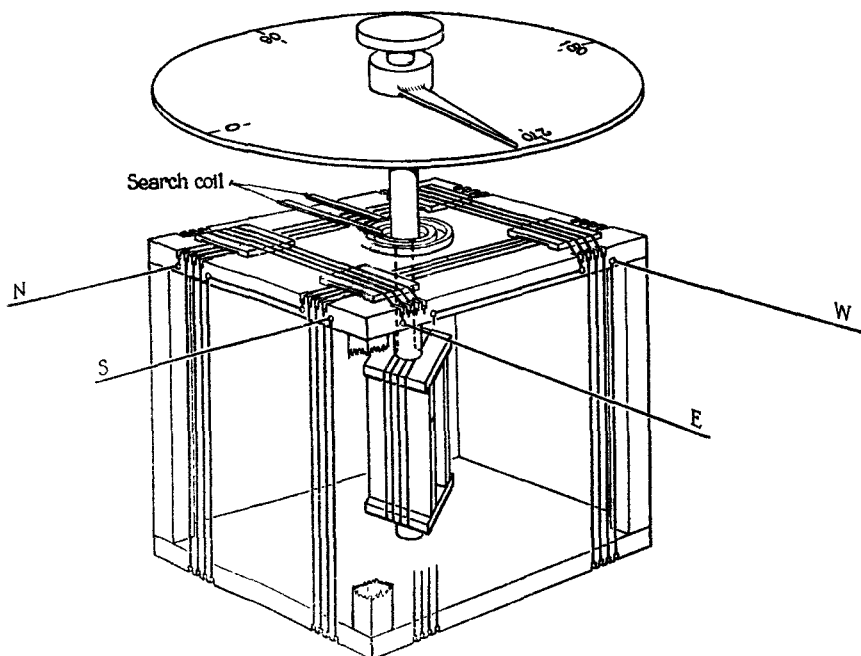


FIG. 27, CHAP. XVI — Radio-goniometer.

54. The aerials of the Bellini-Tosi direction-finder consist of two vertical loops of wire. As a rule one is erected with its plane in the North and South (true) meridian, and the other perpendicular thereto, about a common axis. They may be of any shape and size, but are often triangular in order that they may be supported by a single tall mast together with four shorter ones. For theoretical purposes we may suppose them to be of equal size and rectangular in shape, the connections to the radio-goniometer being shown diagrammatically in fig. 28. The aerial N S is connected to one fixed coil N'S' and the aerial E W to the other fixed coil E'W'.

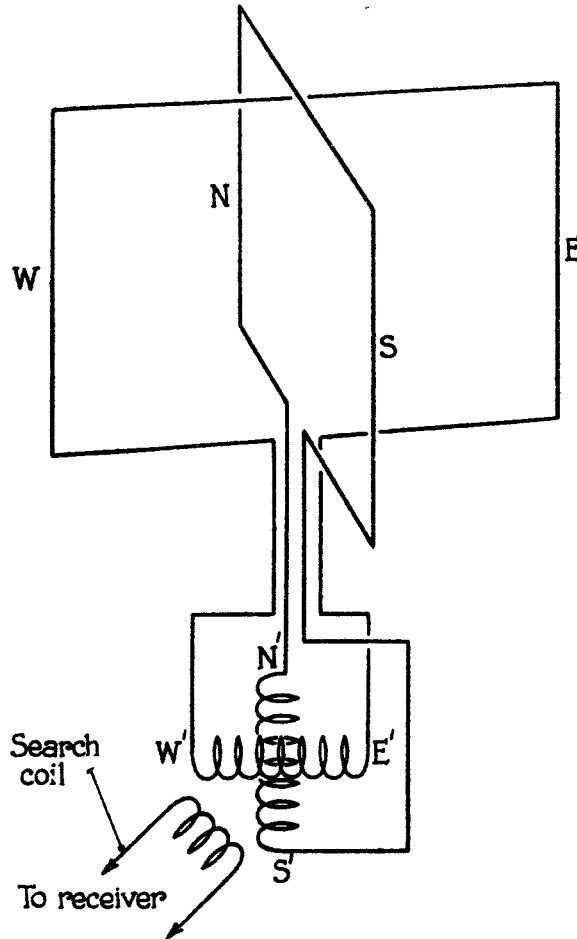


FIG. 28, CHAP. XVI.—Bellini-Tosi D/F.

As already stated the search coil is connected to the W/T receiver, suitable tuned, coupled circuits being generally employed. The loop aeralis and radio-goniometer are shown in sectional plan in fig. 29. Suppose an electro-magnetic wave to arrive at the aerial system from a transmitter situated due north (fig. 29a); a resultant E.M.F. will then be set up in the loop N S. A corresponding current will flow in the loop itself and in the fixed coil N'S' connected to it, setting up an alternating flux in and around the coil. The directions of the current and flux at a particular instant are indicated in the conventional manner in the diagram. No E.M.F. will however be set up in the loop E W, and the fixed coil E' W' establishes no flux. If the search coil is

CHAPTER XVI.—PARA. 55

rotated, therefore, the signal strength will be a maximum when it is in the position shown, for the flux-linkage is then maximum. On the other hand, if the axis of the search coil is perpendicular to the flux, the induced E.M.F. will be zero.

55. Fig. 29b shows the conditions which arise when the transmitter is situated to the north-east. The wave now reaches the sides N and E of the two loops simultaneously, and after a very short interval, the sides S and W. The direction of current in each loop circuit at a given instant is again indicated conventionally. As the two circuits are assumed to be electrically similar,

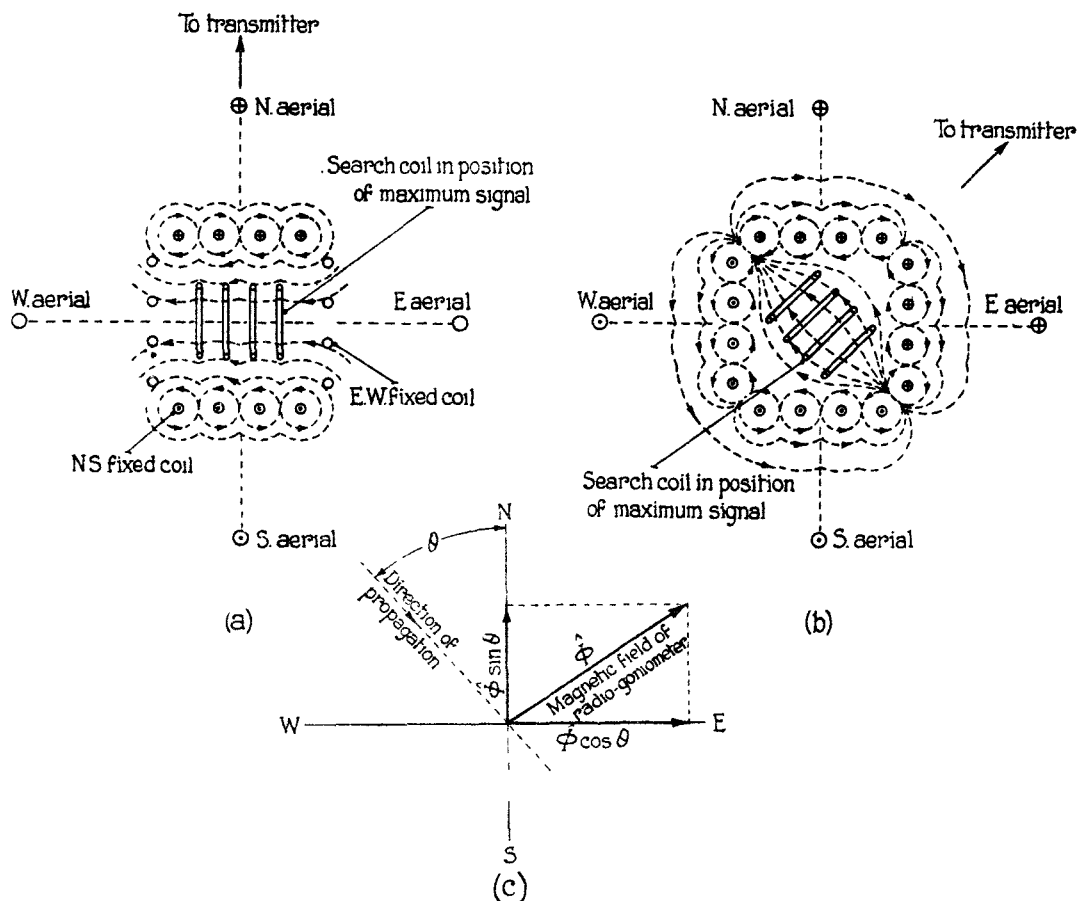


FIG. 29, CHAP. XVI.—Action of radio-goniometer.

the currents in the coils N' S', E' W' are equal and give rise to equal magnetic fluxes. These fluxes combine in such a manner that the resultant flux is in a diagonal direction. In the diagram, the search coil is in the position of maximum flux-linkage. In both the cases illustrated, then, a pointer mounted on the spindle of the search coil, parallel to the turns of the winding, would indicate the direction of the transmitter by the maximum method. It will also be noted that provided the connections between loops and fixed coils are correctly made, and a bearing plate rigidly attached to the fixed coils, the radio-goniometer itself may be placed in any convenient position without affecting the accuracy.

56. Having dealt with the principle physically, let us suppose a vertically polarized electric field $\gamma = \hat{r} \sin \omega t$ to be incident upon the aerial system at an angle θ , reckoned from true north. From paragraph 6, the resultant E.M.F. in the loop N S is

$$e_{NS} = \frac{2\pi A \hat{r}}{\lambda} \cos \theta \cos \omega t$$

and in the loop E W

$$e_{EW} = \frac{2\pi A \hat{r}}{\lambda} \sin \theta \cos \omega t.$$

Let each loop and search coil have a total impedance Z ; the currents will then be

$$i_{NS} = \frac{e_{NS}}{Z}$$

$$i_{EW} = \frac{e_{EW}}{Z}.$$

For simplicity, we may assume that the resistance of each circuit is negligible compared to its inductive reactance. We then have

$$i_{NS} = \frac{2\pi A \hat{r}}{\lambda \omega L} \cos \theta \cos \left(\omega t - \frac{\pi}{2} \right)$$

since

$$\lambda = \frac{c}{f} = \frac{2\pi c}{\omega}$$

$$i_{NS} = \frac{A \hat{r}}{cL} \cos \theta \cos \left(\omega t - \frac{\pi}{2} \right)$$

and similarly

$$i_{EW} = \frac{A \hat{r}}{cL} \sin \theta \cos \left(\omega t - \frac{\pi}{2} \right).$$

The magnitude of the current in each fixed coil is therefore independent of frequency. Each current will set up a flux proportional to the current and to the number of turns on the fixed coil. These fluxes may be written

$$\varphi_{NS} = \hat{\Phi} \cos \theta \cos \left(\omega t - \frac{\pi}{2} \right)$$

$$\varphi_{EW} = \hat{\Phi} \sin \theta \cos \left(\omega t - \frac{\pi}{2} \right)$$

and are shown vectorially in fig. 29c. The magnitude of the resultant flux is

$$\sqrt{[\hat{\Phi}^2 \cos^2 \theta + \hat{\Phi}^2 \sin^2 \theta]} = \hat{\Phi},$$

and is therefore independent of the direction of incidence of the wave. The direction of this flux is such that it makes an angle θ with a plane perpendicular to the turns of the fixed coil N'S', and maximum E.M.F. will be induced in the search coil when it links with the greatest flux. On the other hand, when the search coil is so orientated that it links with no flux at all, no E.M.F. will be induced in it, and therefore, as the coil is rotated, the signals in the telephone receivers are found to indicate two maxima and two minima, spaced alternately 90° apart, just as is found with the rotating frame aerial. To a near approximation, the polar diagram of the arrangement is a figure-of-eight pattern, and in practice, the bearing is always obtained by observation of the two minima.

57. The earliest form of the Bellini-Tosi system made use of two vertical loops each of which was separately tuned to the desired signal, and the search coil was small compared to the fixed coils, in order that it should rotate in a very uniform portion of the field. This arrangement

CHAPTER XVI.—PARAS. 58-60

was very difficult to maintain in working order and was superseded by the aperiodic loop system which is substantially that described in the preceeding paragraph. In this form of apparatus the radio-goniometer is so constructed that the coupling between the search and fixed coils is as tight as possible. The search coil then forms part of an intermediate or link circuit between the fixed coils and the tuned circuit forming the input to the receiver proper. The coupling at the latter point is usually by means of a shielded radio-frequency transformer (*see* paragraph 20) in order to prevent electro-static coupling between the receiver and the aerials and so eliminate any vertical error due to the receiver itself. The link circuit is usually aperiodic but in some installations it is tuned. The tightly-coupled search coil gives rise to an error due to the fact that the field in which it rotates is far from uniform. The error is a maximum (of the order of 2°) at about 22.5° , 67.5° , 112.5° and so on ; as there are eight points of maximum error in the 360° it is called octantal error. It is practically eliminated by suitable design of the search coil, one method being to place a second winding on the search coil in series with the first, the two windings making an angle of approximately 45° .

58. Apart from errors due to the development of a faulty connection or the like, the most likely errors to be found in the Bellini-Tosi aperiodic aerial system are :

- (i) mutual inductance between aerials,
- (ii) vertical error.

The presence of mutual inductance between the aerials may be detected by disconnecting one aerial from the radio-goniometer and energizing it by coupling to a perfectly-screened oscillator. If the aerials are truly perpendicular to each other, the oscillator radiation will not be detectable in the receiver, which is still coupled to the unenergized aerial by the radio-goniometer. Once the aerials are correctly set up, it is most important that their position should be maintained, and it is necessary to pull the wires very taut during erection in order to prevent swaying. Each aerial is tested independently for vertical error by disconnecting the other from the radio-goniometer and then observing whether the two minima of a distant transmitter are 180° apart, as well as reasonably sharp. The two sides of each aerial must be perfectly symmetrical with respect to earth. The lower, approximately horizontal portions are generally to be first suspected, in the event of vertical error due to aerial asymmetry, because these have the greatest capacitance to earth. Instrumental vertical is eliminated by the employment of shielded transformers for coupling purposes, and by earthing the midpoint of each fixed coil of the radio-goniometer (*cf.* paragraph 19).

Comparison between rotating loop and Bellini-Tosi systems

59. It must be understood that since the fixed aerials and radio-goniometer of the Bellini-Tosi system are the electrical equivalent of a rotating frame aerial, there is no question of one system giving more accurate bearings than the other. Provided it can be mounted in a suitable position, therefore, the rotating coil has the advantage of simplicity. An electrical defect, such as a badly soldered connection, may result in weak signals, or even to no signals at all, but will rarely give rise to an erroneous bearing. In the Bellini-Tosi system, the two aerial circuits must be perfectly balanced, and the sudden occurrence of a fault in one circuit may cause an error of 90° in the apparent bearing. The Bellini-Tosi type of D/F aerial has, in recent years, been largely supplanted by spaced open aerials, in order to reduce errors due to abnormally polarized waves.

POLARIZATION ERROR

60. Hitherto, it has been assumed that the electro-magnetic wave is incident upon the D/F aerial with normal polarization. If the wave travels for an appreciable distance over the surface of the earth before reaching the receiving aerial, this must be the case (Chapter XIV). The incident wave may however be abnormally polarized (i) when it is transmitted from an aeroplane which is so near to the receiving aerial that an appreciable amount of energy is incident at an

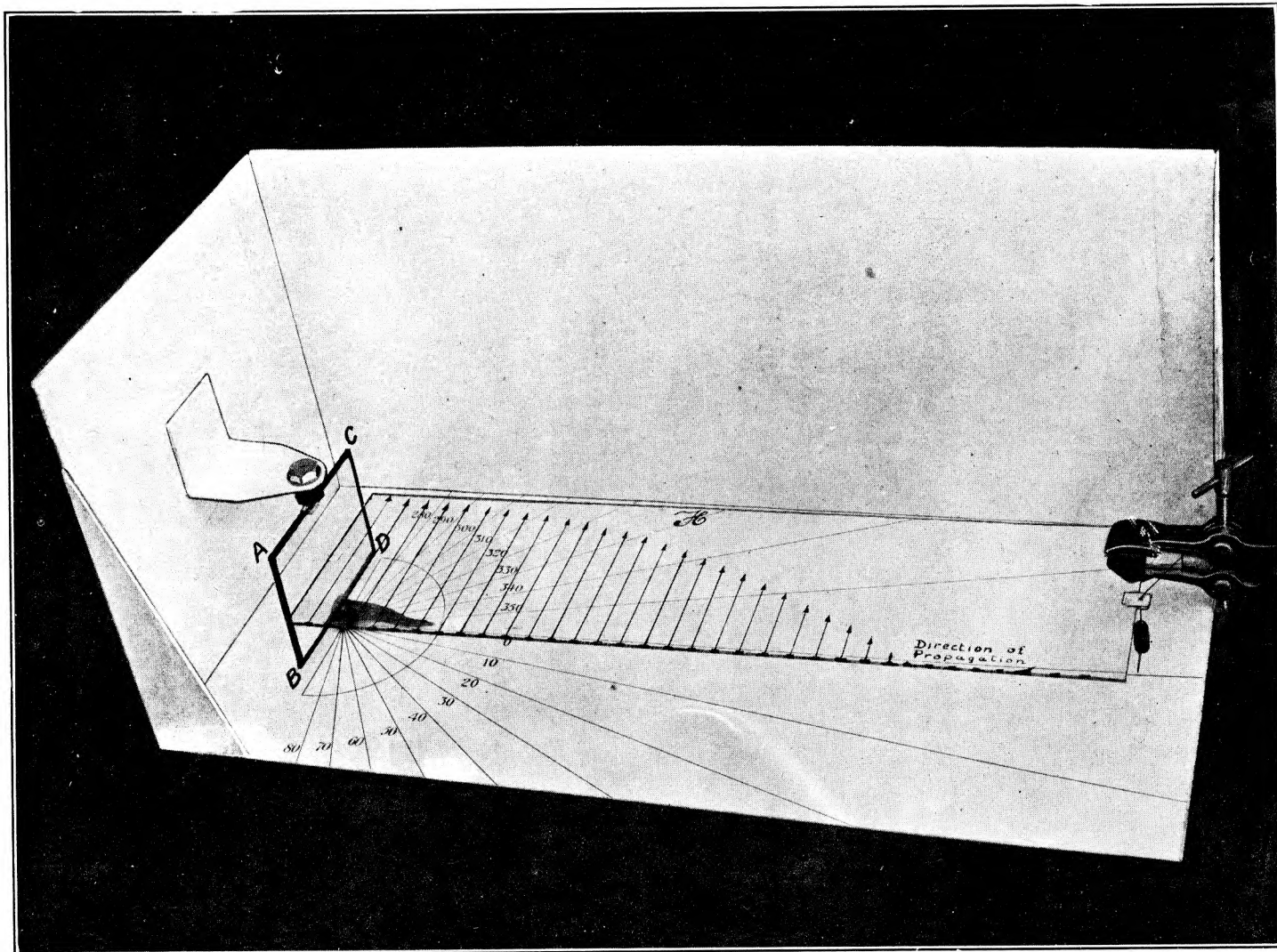


FIG. 30, CHAP. XVI.—Minimum when wave is normally polarized

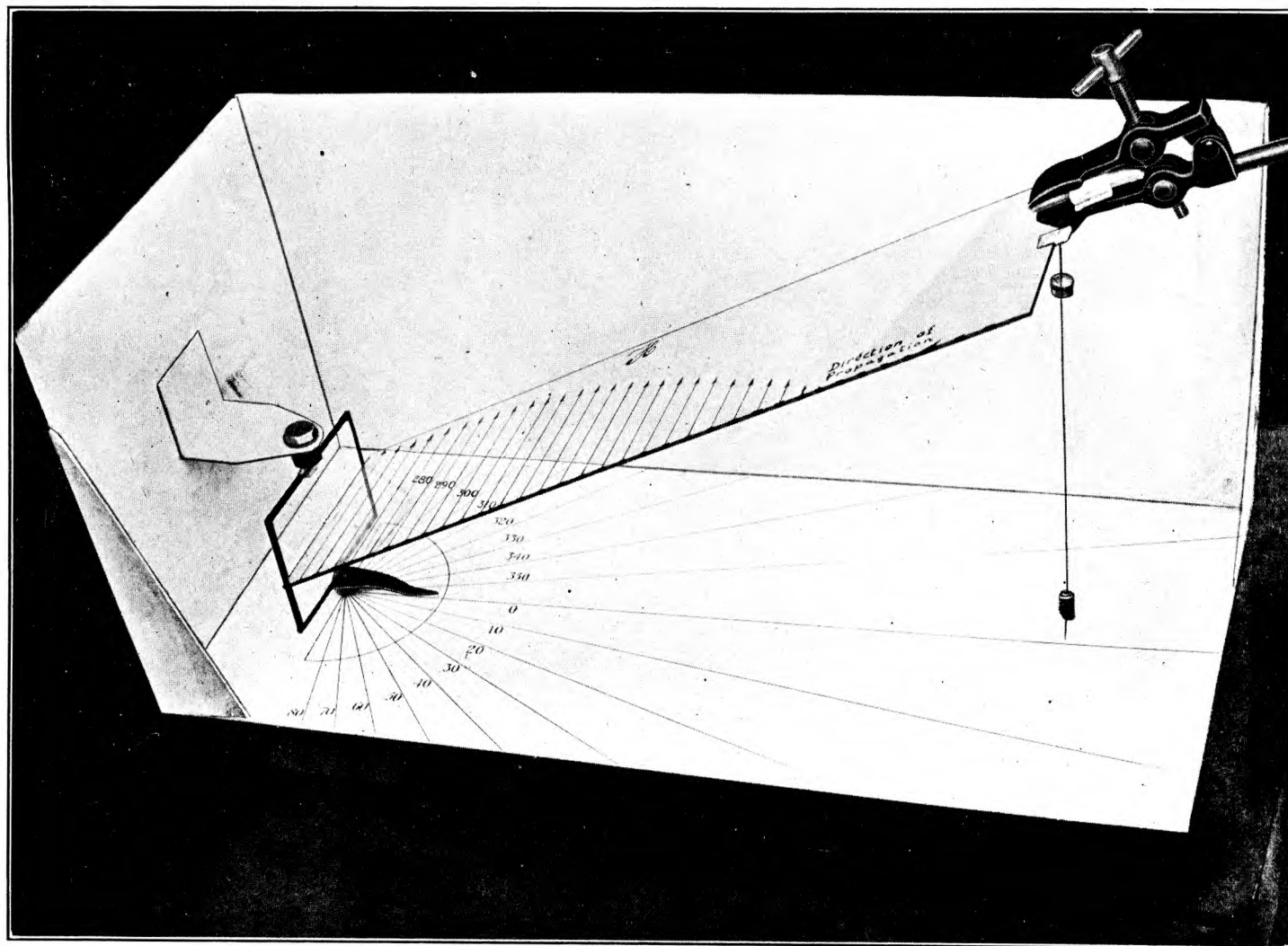


FIG. 31, CHAP. XVI.—Minimum with normally polarized down-coming wave.

angle to the horizontal or (ii) when it reaches the receiver by reflection from the ionosphere. The effects are the same in either case and are as follow :—

- (a) The minima may be blurred.
- (b) The minima may be sharp, but successive bearings of a stationary transmitter are not in agreement. The apparent bearing may vary during the time it is being obtained.
- (c) It may be impossible to observe any minima.

In the following discussion, abnormally polarized waves are discussed with reference to a rotating loop aerial on the ground, but the deductions are equally applicable to rotating loop aerials in aircraft and to fixed loop aerials used in conjunction with a radio-goniometer.

61. The action of waves which are incident upon the loop aerial at an angle to the horizontal—which may for brevity be called “downcoming waves”—is more easily understood if we consider the wave to induce an E.M.F. by virtue of the velocity of its magnetic field, and may be illustrated in a conventional manner by the model shown in succeeding figures. In this model, the loop aerial has vertical sides A B, C D, and horizontal sides A C, B D; it is capable of rotation about a vertical axis and is mounted over a bearing plate. For simplicity the transmitter is supposed to be situated due north (0°) of the loop. A pointer is fitted indicating the apparent bearing of the transmitter (or its reciprocal) when the loop is in the position giving the minimum signal.

62. In the first diagram, fig. 30, the wave is arriving in parallelism with the ground, and is normally polarized. The magnetic flux of the wave will induce an E.M.F. in each of the vertical wires A B, C D, but not in the horizontal wires A C, B D, because these are parallel to the flux. For simplicity we may confine our attention to the peak value of the flux; if the loop is turned so that the pointer indicates, e.g. 60° , the peak value of the flux will reach the wire C D before it reaches the wire A B, and similarly for all other instantaneous values, i.e. a sinusoidal E.M.F. will be induced in each wire, but that in A B will lag behind C D. The resultant E.M.F. acting round the loop is the vector difference of these. If, however, the loop is orientated as shown in the figure, the peak value of magnetic flux links with both vertical wires simultaneously, and this is true of all other instantaneous values, hence the induced E.M.F. in A B is in phase with that in C D and the resultant E.M.F. is zero. The plane of the loop is perpendicular to the direction of the transmitter and the pointer correctly indicates the direction of the transmitter, with an ambiguity of 180° . Now consider the conditions which arise in the reception of a downcoming wave which is normally polarized, and illustrated by the model in fig. 31. The magnetic field again links only with the vertical sides of the loop, and the induced E.M.F. in A B will be in phase with that in C D when the plane of the loop is perpendicular to the bearing of the transmitter. The pointer, then, gives the correct bearing of a downcoming wave, provided it is normally polarized.

63. Since normal polarization is generally attributed to the fact that the wave is travelling over the surface of a conductive earth, there is no reason why downcoming waves should be normally polarized, whereas there are many phenomena which may cause the polarization to be abnormal, even if the wave leaves the transmitter with normal polarization. Suppose the wave to reach the loop aerial in a downward direction, and to be polarized at an angle of 90° , so that the magnetic field is in the vertical plane and the electric field horizontal. Let the loop be turned so that its plane is perpendicular to the direction, in azimuth, of the transmitter, so that the pointer gives the correct bearing. The magnetic field does not link with the vertical wires A B, C D, but links with the horizontal wires A C, B D, and although no E.M.F. is induced in the vertical wires, there is a resultant E.M.F. acting round the loop in this position. Suppose we now rotate the loop until a minimum signal is obtained; this will occur when the loop is in the position shown in fig. 32, for in these conditions the magnetic flux does not link with any of the four wires constituting the loop. The pointer indicates the apparent bearing as 270° or 90° , i.e. there is an error of 90° in the bearing given by the minimum method.

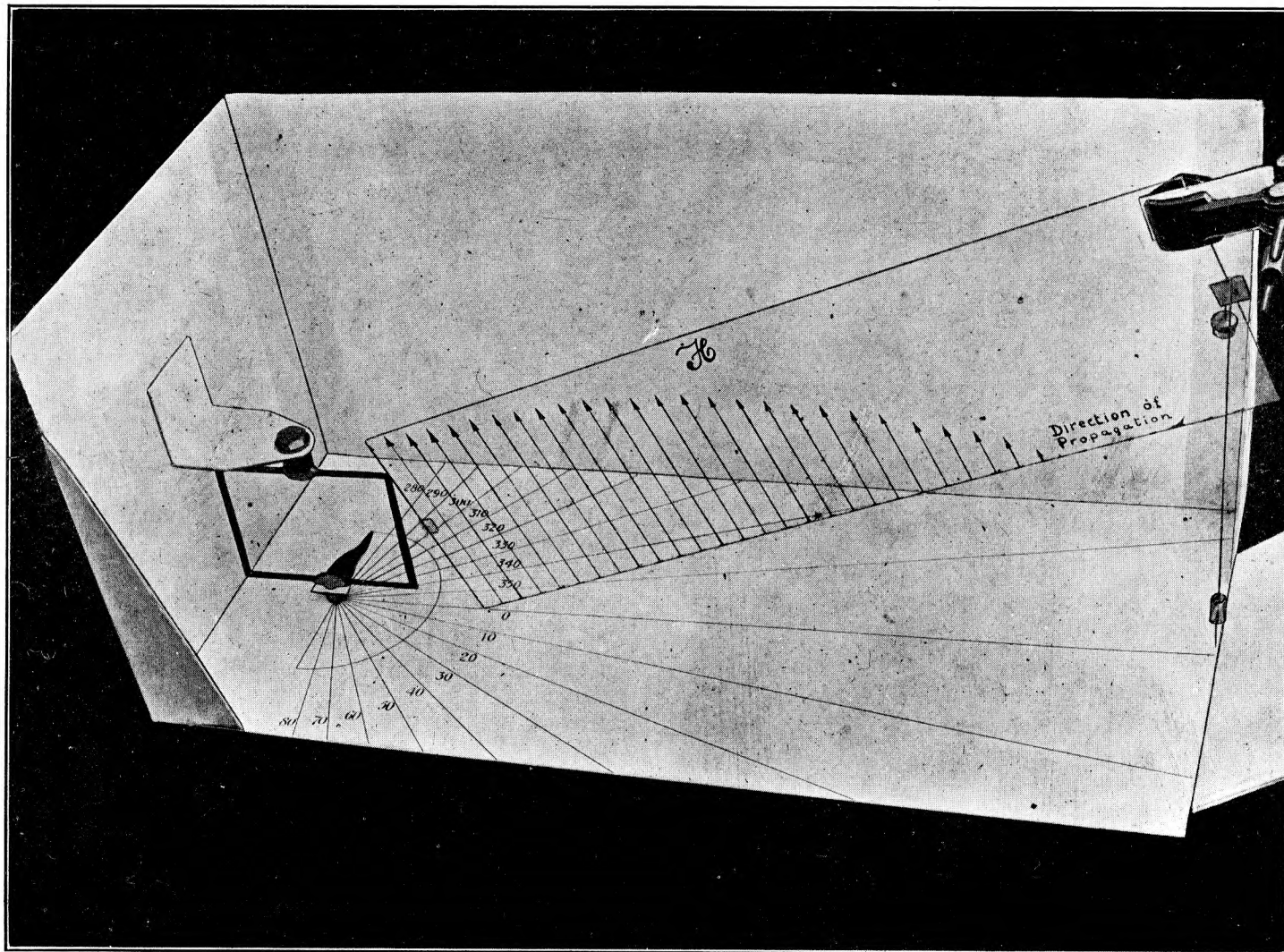


FIG. 32, CHAP. XVI.—Minimum with horizontally polarized wave.

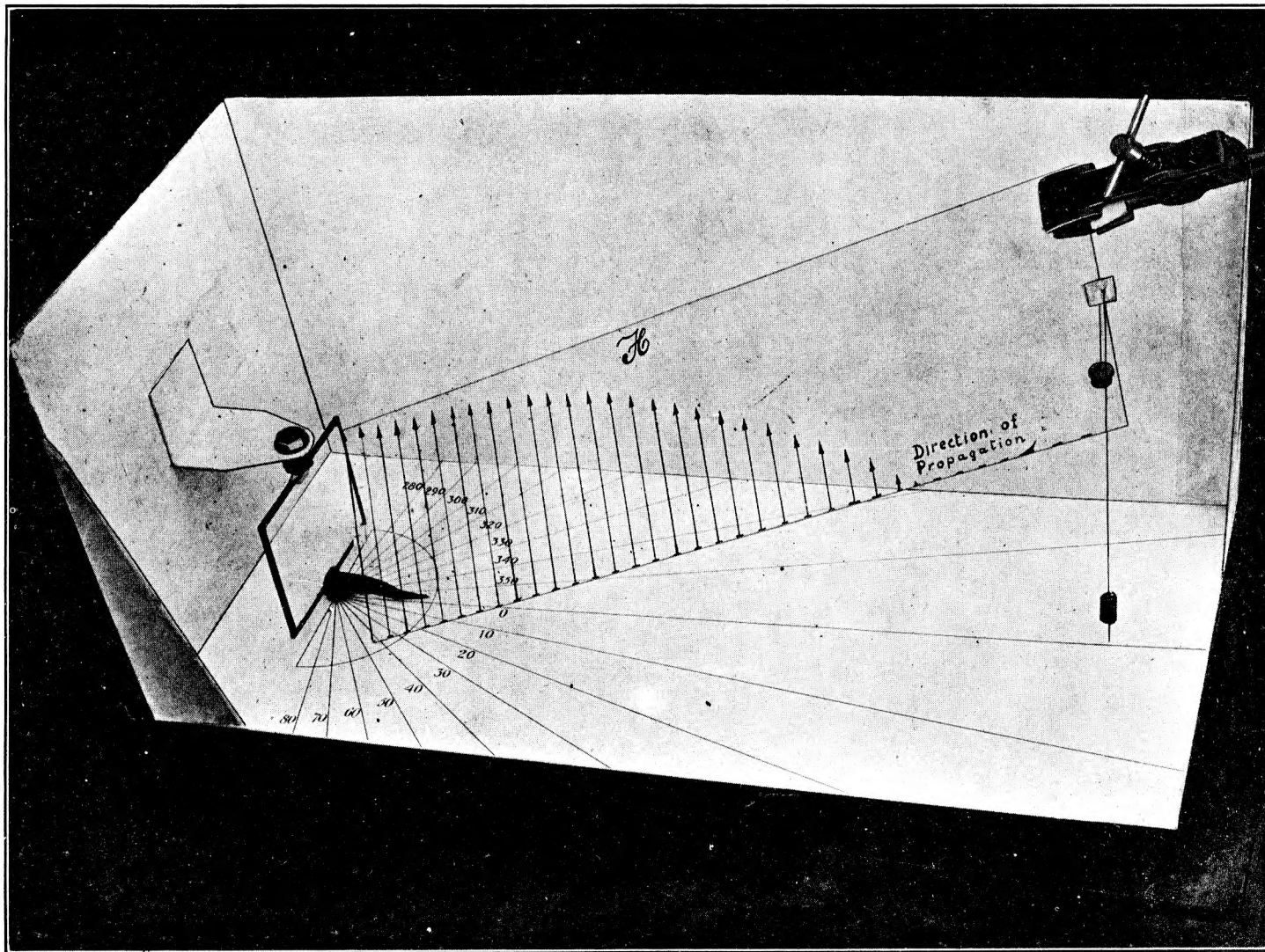


FIG. 33, CHAP. XVI.—Down-coming, abnormally polarized wave.

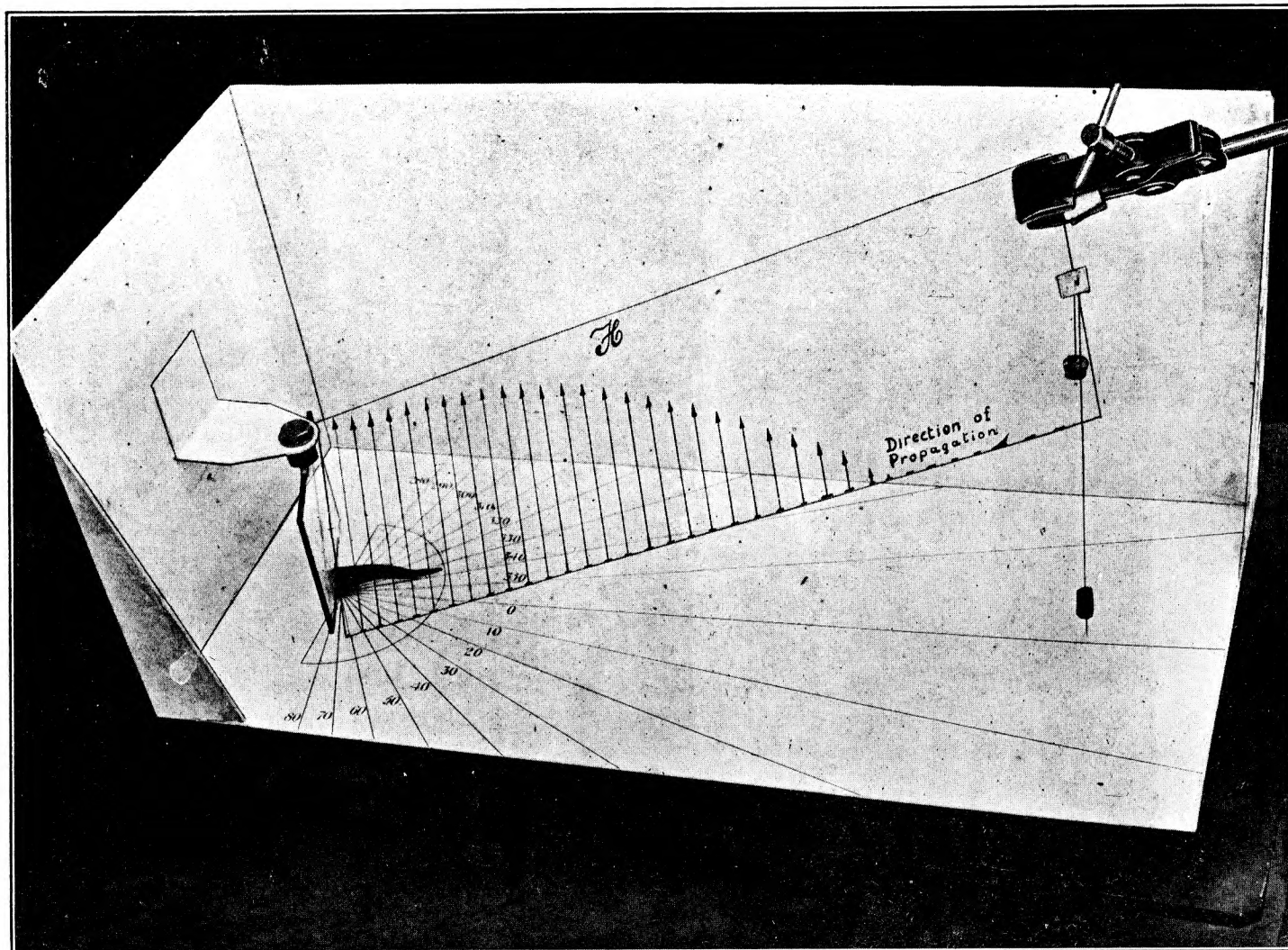


FIG. 34, CHAP. XVI.—Minimum with normally polarized, down-coming wave.

64. The most general case is illustrated in figs. 33 and 34. Here a downcoming wave is polarized at an angle of about 50° , and at the moment illustrated in fig. 33, the vertical wire C D is linking with the peak value of the magnetic flux; the wire A B is however linking with a flux of smaller magnitude. In addition, both the upper and lower horizontal sides now link with the flux; the E.M.F.s in these sides are not in phase with each other, for any given instantaneous value of flux links with the upper before linking with the lower wire. When the plane of the loop is perpendicular to the bearing of the transmitter, therefore, there is a resultant E.M.F. acting round the loop and minimum signal is not obtained in this position. If however the loop is rotated a position will be found (fig. 34), in which the peak magnetic flux links with all four sides simultaneously; the two E.M.F.s in the vertical sides A B and C D are then equal in amplitude and in phase, giving zero resultant, and the E.M.F.s in the horizontal sides A C and B D are also equal in amplitude and in phase, thus also giving zero resultant, consequently the minimum signal is obtained with the loop in this position. In fig. 34 the apparent bearing of the transmitter, as found by the minimum method, is about 335° i.e. the bearing is 25° in error.

65. The foregoing photographs therefore demonstrate that if a loop aerial is rotatable about a vertical axis, it will give minimum signal when the plane of the loop is perpendicular to the position line upon which the transmitter lies, only if the wave is polarized with its magnetic field in the horizontal plane; if the wave is polarized in any other plane, and this polarization and the angle of incidence remain constant, the loop will give sharp minima, but they will be displaced to any extent up to 90° . Since however in the case of downcoming signals there is no agency tending to maintain constant polarization, we find that plane polarized waves of this kind give minima which move round on the bearing plate, apparently indicating that the bearing of the transmitter is changing from instant to instant. Actually, of course, what is changing is the angle of polarization and possibly the angle of incidence also. It has also been tacitly assumed that the downcoming wave reaches the loop by only a single path. This is unlikely to be the case and therefore the bearing giving minimum signal for one path and angle of polarization, will give an appreciable signal owing to the energy arriving by some other path. Instead of sharp though displaced minima, we shall obtain blurred and moving minima. This is in fact the most frequent indication that abnormally polarized signals are being received. If the wave is circularly or elliptically polarized, no orientation of the loop will give a minimum signal on a constant bearing. Again, interference between waves travelling along different paths may cause fading, which may be mistaken for a minimum and give rise to an erroneous bearing.

Adcock aerial systems

66. The effect of abnormally polarized waves upon the horizontal members of a loop aerial may be eliminated by the virtual removal of these members; this modification was first proposed by Lieutenant F. Adcock, R.E., in 1919, but its practical development occupied many years. The Adcock aerial system may be adapted for use either as the equivalent of a rotating loop, or in conjunction with a radio-goniometer. The latter form of installation is referred to as the Bellini-Tosi-Adcock direction-finder.

U type

67. The simplest form of Adcock aerial is that called the U type. This consists of two vertical aeriels, fig. 35a, the lower ends of which are connected by a horizontal wire to the inductance L. The latter serves to couple the aerial to the receiver in the usual manner. If such an aerial were erected over perfectly conductive ground, with its horizontal member at or very near ground level, a horizontally polarized wave would have no effect whatever upon the system, but it would be acted upon by vertically polarized waves like an ordinary loop. Under certain conditions, e.g. over sea water or soil of very high conductivity, the polarization error of such an aerial may be less than that of the loop, but under ordinary conditions its performance is little better. An improvement is effected by enclosing the lower horizontal member in a screening tube earthed at the outer ends; this type is referred to as a "screened U" (fig. 35b).

H type

68. In the elevated or H type, the vertical aerials are in the form of dipoles, the lower ends of which are clear of the ground. The horizontal leads connecting opposite members are situated at the midpoint of the dipoles and are therefore at a height above ground at least equal to one-half the total height of the aerials (fig. 35c). This system is entirely free from polarization error if it is sufficiently remote from the earth's surface, but the proximity of the ground introduces a certain degree of asymmetry, which in turn gives rise to polarization error. The reason for this is easily seen from the fact that the capacitance to earth of the lower half of each dipole is greater than that of the upper half. Unfortunately however it is not possible to compensate for this simply by shortening the lower members, for the effect of a given degree of asymmetry

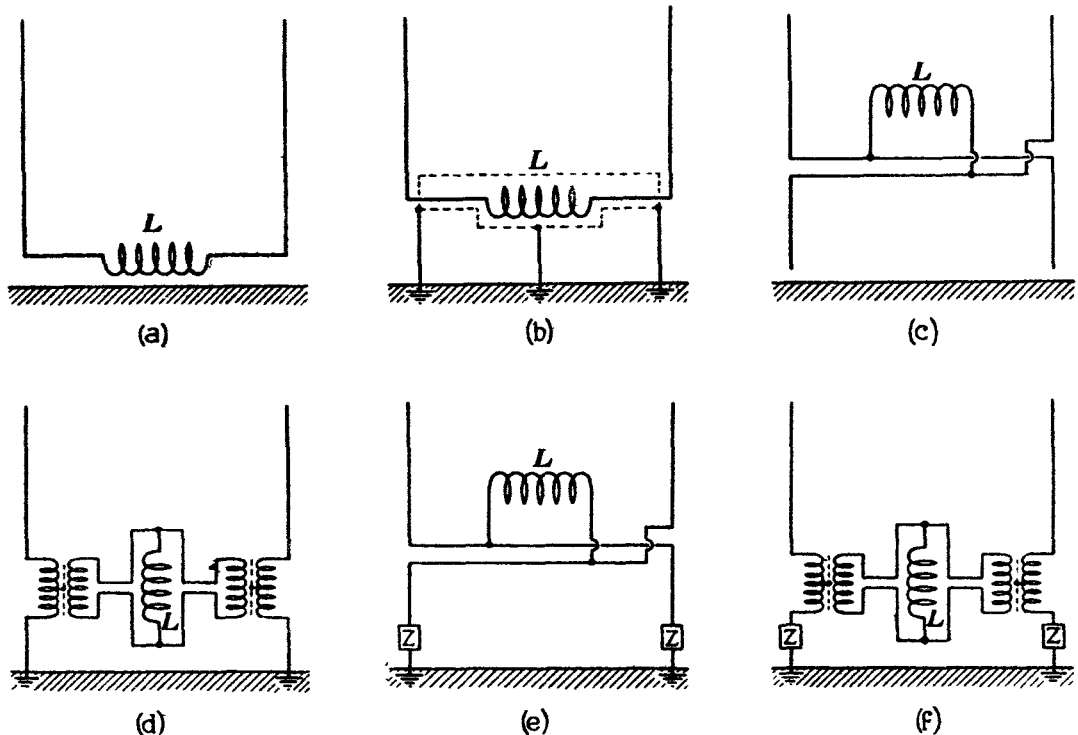


FIG. 35, CHAP. XVI.—Types of Adcock aerial.

varies with the angle of incidence of the downcoming, abnormally polarized wave. An advantage of the H type is that the polarization error is practically independent of the electrical nature of the ground.

Coupled type

69. The coupled type, shown in fig. 35d, is distinguished by the fact that there is no direct connection between the two spaced aerials forming the directional pair. This form may be made to give a very good performance on medium frequencies and is superior to those previously mentioned on high frequencies. The chief factor which allows the abnormally polarized wave to affect the receiver is the capacitance coupling between the windings of the coupling transformers and the latter are usually of the shielded type (paragraph 20). The coupled aerial is obviously ill-adapted for use as a rotating system, but there is no difficulty in using two pairs of aerials in conjunction with a radio-goniometer.

Balanced type

70. The balanced type, fig. 35e, is practically the same as the H system, except that the lower members of the aerials are shorter than the upper members, and are connected to earth through suitable impedances, which are denoted by Z in the diagram. The polarization error can be reduced to a very small magnitude by adjustment of these, but for complete elimination it would be necessary to design the system for operation on a single frequency.

Balanced coupled type

71. Fig. 35f shows a combination of the "balanced" and "coupled" systems known as the balanced-coupled system. It appears that this arrangement may be made almost free from polarization error. The performance depends to a great extent upon the correct design of the coupling transformers, and also upon the arrangement of the feeders. In this respect it is to be noted that although it would appear desirable to use buried, screened feeders, it may be found preferable to use overhead quadruple wires. This form of line is illustrated in fig. 36. Each diagonally opposite pair of wires is linked together at the ends to form one of a pair of twin transmission lines, the arrangement being practically non-radiative and therefore having negligible pick-up properties.

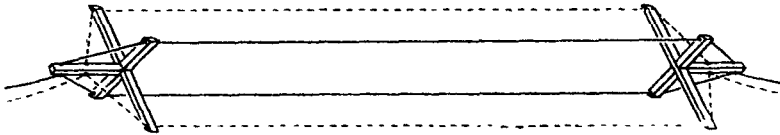


FIG. 36; CHAP. XVI.—Twin feeder for Adcock aerials.

72. In the design of an Adcock aerial the following points must be taken into consideration. First, in a rotating aerial system, it is not possible to increase the pick-up factor by increasing the number of vertical sides as in a frame aerial; increased sensitivity can only be obtained by an increase in the area embraced by the aerial system. Second, as each aerial is an open oscillator and its height is only a fraction of a wavelength, its effective height is of the order of only one-half the actual height. Third, each pair of aerials, considered as a single unit, offers a large capacitive reactance, whereas each turn of a closed loop offers a comparatively small inductive reactance. In certain cases, the aerial capacitance is increased by using multiple wire vertical aerials instead of single wires. This also reduces the R.F. resistance and so increases the pick-up factor. The Adcock aerial system is usually coupled to a closed oscillatory circuit which is so adjusted that the P.D. across the condenser terminals is a maximum. The pick-up factor is then the ratio of the condenser P.D. to the field strength. It is found that to obtain a high pick-up factor the area embraced by the aerials should be as large as possible, subject to the limitation imposed by the necessity for keeping the fundamental frequency above the highest operating frequency. In the Bellini-Tosi-Adcock direction-finder a further limitation is imposed by the fact that if the spacing, d , between the two aerials of a pair is an appreciable fraction of a wavelength, the approximation $\sin \frac{\pi d}{\lambda} = \frac{\pi d}{\lambda}$ does not hold. The result is that an octantal error is introduced. This error is negligible if the ratio $\frac{d}{\lambda}$ is less than 0.1, and of the order of 1° if $\frac{d}{\lambda} = 0.2$.

Selection of site for ground D/F station

73. The site for a ground D/F station should, if possible, be a flat open piece of land about half a mile in diameter, of uniform soil surface, and not near the sea. It need not be absolutely bare, rough grassland and hedges do not matter, but the site should be free from trees, buildings, streams and other objects specifically dealt with later. A site fulfilling these conditions will be practically perfect, but as a rule it will be necessary to accept one which falls short of this ideal.

CHAPTER XVI.—PARAS. 74-77

The whole area under consideration may be hilly, woody, industrialized or near the sea ; it is also desirable to locate the D/F station conveniently near to the remainder of the station buildings. The under-mentioned factors will contribute to the site errors.

- (i) Trees and woods.
- (ii) Cables and wires laid along or under the ground.
- (iii) Overhead cables and wires.
- (iv) Mountains, hills and cliffs.
- (v) Buildings.
- (vi) Roads and railways.
- (vii) Rivers and streams.
- (viii) Aerials and vertical conductors such as metallic poles and pylons.
- (ix) Wire fences.
- (x) Coastal refraction.

The following discussion on the effect of these positions on the performance of the direction finder covers the frequency range 100 to 15,000 kc/s.

74. Trees and woods. Large masses of trees within half a mile will probably produce considerable errors, which amount to 5 or 6 degrees. This type of error depends upon the frequency, but the rate of variation with frequency is uncertain. It must be realized that errors of such magnitude do not occur on every wooded site. It might, for example, happen that a degree of symmetry in the position of the trees around the site would render the error negligible. An error may be produced by a single tree, and a direction finder located within several yards of a tall tree may be expected to experience an error of several degrees on short waves. For example, an error exceeding nine degrees has been observed at a distance of ten yards from a tree 60 feet high ; a distance of 30 yards from a single tree, however, may be considered great enough to reduce all errors to zero.

75. Cables and wires. It appears safe to conclude that cables and wires may be led into a D/F operating room along the ground, provided they are earthed at the point where they enter and at points less than $\frac{\lambda}{2}$ apart for a distance of 200 yards or more. In the case of screened cable, earthing implies connection of the metallic sheath to earth and in the case of an unscreened pair of wires a radio-frequency earth for each wire by means of suitable condensers. Wires should not be run on overhead poles within about 200 yards of the site, in the case of medium frequencies. For frequencies of the order of 15,000 kc/s. it may be necessary to bring the wires down 400 yards away or even more.

76. Overhead cables and wires. A certain amount of experimental work on long waves with small loops shows that a distance of 200 yards, from telegraph wires 10 yards high, is sufficient to reduce all errors to a negligible quantity. Overhead wires carried by tall pylons may be expected to have greater influence than telegraph wires but it is unlikely that errors will be experienced at distances greater than 400 yards. The distortion produced by the pylons themselves must also be taken into account. These constitute radiating vertical aerials. In general it may be assumed that a D/F receiver erected within 50 yards of overhead telegraph lines or within 100 yards of a high power line will be subject to very grave site error which might be as much as 10°, while at a distance of 200 yards from a high power line the error might amount to 2° or 3°. At 400 yards from high power lines the site error would probably be negligible under all circumstances.

77. Mountains, hills and cliffs. A site in a mountainous country is liable to a large error (exceeding 10°) as a result of reflection and bending of the waves. The same applies to hills but to a lesser degree ; in hilly country it is preferable to choose a site on high rather than low ground, but a sloping hillside site should be avoided. Any sudden change of slope is liable to give errors

up to 10° on short waves, for instance, in the vicinity of a cliff, complete immunity will not be obtained unless the site is at least 10λ from the edge of the cliff. On medium frequencies a distance of 200 yards will probably ensure that the error is less than 3° . The effect produced by mountain ranges may persist into low-lying flat country situated well away from the mountains.

78. Buildings. A metal building is liable to cause more disturbance than one made of stone, brick or other non-conductive material, but as most large buildings contain metal work in the form of gas pipes, electric light wires, water systems, etc., it is probably desirable to consider all buildings on the same footing. If the building as a whole, or any part of it, resonates to the frequency in use, the disturbing effect may be large at a distance of many wavelengths from the building. Thus a large modern type of building may be expected to give errors of several degrees on a direction finder erected within 100 yards but would probably have negligible effect at a distance of 200 yards, except for the above-mentioned resonant effect which however would probably be negligible at distances of 500 yards or over.

79. Roads and railways. Ordinary stone metalled roads are not likely to give rise to site error at any frequency, but this is not the case with a modern type of concrete road which probably contains steel in the foundation. However, it is considered that roads and railways are unlikely to produce errors exceeding 5° and if over 100 yards from the site will produce no appreciable error. Electric railways may cause considerable interference with reception at greater distances than this.

80. Rivers, streams and canals. A small stream or brook is electrically of similar nature to the reinforced concrete road mentioned above but its conductivity is of course much smaller, and it will probably have negligible effect if it is more than 30 yards from the direction finder. A large river will produce a greater effect and it is considered undesirable to site a station within about 500 yards.

81. Aerials and vertical conductors. An aerial or vertical conductor may effect a direction-finder in two ways, which may be distinguished by the terms "re-radiation error" and "screening error" respectively. The first is most likely to be caused by resonating conductors at distances from the receiver large compared with the spacing of the aerial and represents a genuine re-radiated wave. The second occurs chiefly when untuned aerials or resonating vertical conductors are situated at distances comparatively near to the direction-finder. This is in the nature of a screening effect, which, acting unequally on each of the aerials, causes an unbalancing of the aerial system similar to that caused by inequality of their effective heights. To avoid these effects no vertical conductor of height comparable with that of the D/F aerial should be allowed within a distance of three times the aerial spacing, and no tuned aerial or vertical conductors in the D/F waveband should exist within six wavelengths of the D/F receiver. Lamp standards, lightning conductors and pylons all come within this category.

82. Wire fences. Experiments performed with a small fence in the neighbourhood of a short-wave Adcock direction-finder indicates that an insulated length of wire fence in resonance with the operating wave might produce an appreciable polarization error. The fence when well earthed had no effect. The effect of a long length of wire fencing has not been studied but the experiment makes it probable that the disturbing effect will be considerably less than would be expected and it is anticipated that a fence more than 30 yards away would have no detectable effect. As a matter of precaution however it is advisable to earth all wires on the fence at several points less than $\frac{\lambda}{2}$ apart in the vicinity of the receiver, or preferably to break all wires into sections less than $\frac{\lambda}{2}$ in length by suitable insulators.

83. Coastal errors. Serious errors, i.e., of the order of 5° to 10° , may exist on sites close to the sea unless the wave crosses the coast line at an angle greater than 30° . These errors are greatest for soil of low conductivity and on high frequencies, decreasing as the distance from the coast is increased. Coastal errors are unlikely to be appreciable at distances exceeding 10 miles from the

CHAPTER XVI.—PARAS. 84-85

coast and may be neglected for waves making angles greater than 30° with the coast. There is no method of correcting for coastal error, and it is usual, on charts showing coastal direction-finding stations, to indicate the arc within which accurate bearings cannot be guaranteed.

Radio beacons

84. The simplest form of radio beacon is a transmitting station having an aerial system designed to give substantially uniform radiation in all directions in azimuth. The station will emit a characteristic signal, e.g. its W/T call sign, either continuously or in accordance with a prearranged time schedule, and may therefore be used by aircraft fitted with direction-finding apparatus to obtain a position line, or if sense-finding devices are installed, the bearing of the transmitter with reference to the aircraft. Other forms of beacon, such as the equi-signal and rotating loop types, are designed for the use of aircraft not fitted with D/F apparatus. Equi-signal beacons are primarily intended for homing, and are largely used in America, but not to any extent in this country. One factor which makes their adoption difficult in Europe is the absence of any frequency allocation for this purpose. In the United States the frequency band of 210–410 kc/s. is allotted solely for the air navigation and weather broadcast service.

The loop or frame aerial transmitter

85. In Chapter VII reference is made to the radiating properties of a loop aerial and these may now be discussed in greater detail. In fig. 37 A B and C D represent the two vertical sides of a rectangular loop in elevation (fig. 37a) and in sectional plan (fig. 37b). As in the receiving case, the dimensions h and d are supposed to be small, compared with the wavelength. The lower horizontal wire is assumed to be very near the ground level. If the loop is supplied with a radio-frequency current of constant amplitude in such a manner that the magnitude of the current is the same at all points, all four sides of the loop will give rise to electro-magnetic radiation, the vertical sides setting up a vertically polarized wave and the horizontal sides a horizontally polarized wave. The solid polar diagram of the loop transmitter is therefore rather complex. For simplicity, it is convenient to consider the vertically and horizontally polarized portions of the radiation field separately, and also to postulate a perfectly conductive earth. If the radiation is received at ground level by means of a vertical aerial, the latter will be affected only by the vertically polarized portion of the field.

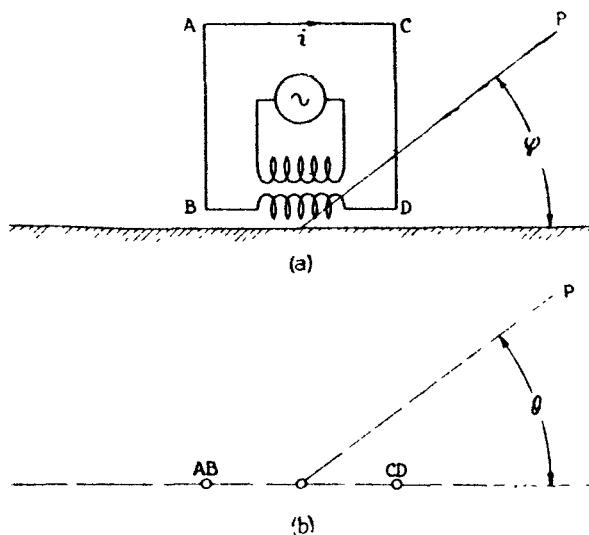


FIG. 37, CHAP. XVI.—Loop aerial transmitter.

86. By methods explained in Chapter XIV it is easily seen that the horizontal polar diagram of this portion is a figure-of-eight pattern. The amplitude of the vertical field strength at any point P, making an angle θ with the line passing through the vertical sides (fig. 37b), is,

$$\begin{aligned}\hat{F}_v &= \frac{240\pi h}{\lambda r} \mathcal{G} \sin \frac{\pi d \cos \theta}{\lambda} \\ &= \frac{240\pi^2 A}{\lambda^2 r} \mathcal{G} \cos \theta,\end{aligned}$$

by making the substitutions $\sin \frac{\pi d}{\lambda} \doteq \frac{\pi d}{\lambda}$, $A = hd$ as in previous paragraphs. The effect of the ground is here taken into account. In free space the field strength would be only one-half of that given. If the point P is situated above ground level, at an angle ϕ as in fig. 37a, the vertical field strength is obtained by multiplying the above expression by a Current Distribution Factor. For the present case (uniform current distribution in the aerial) the C.D. Factor is $\cos \phi$, hence the polar diagram in the plane containing the vertical sides is as shown in fig. 38a. The effect of the finite conductivity and permittivity of the ground is to modify this diagram somewhat as shown in fig. 38b.

87. It is of interest to calculate the R.M.S. current required to obtain a given field strength at any distance from the transmitter, assuming a perfect earth. Let the area of the loop be

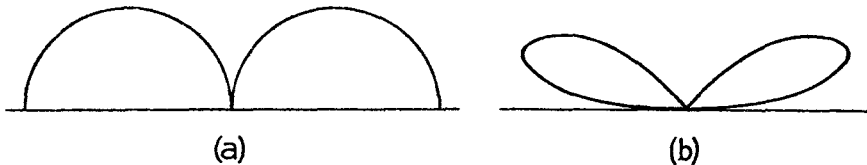


FIG. 38, CHAP. XVI.—Vertical polar diagrams of loop transmitter.

10 square metres, the wavelength 1,000 metres ($f = 300$ kc/s). If the field strength on the ground, at a distance of 100 miles, is to be 50 microvolts per metre, (R.M.S.).

$$F = \frac{149\pi^2 A I}{\lambda^2 r}$$

where r is in miles, hence

$$\begin{aligned}I &= \frac{\lambda^2 r F}{149\pi^2 A} \\ &= \frac{(10^3)^2 \times 100 \times 50}{149 \pi^2 \times 10 \times 1,000} \\ &= 334 \text{ amperes.}\end{aligned}$$

It is difficult to provide such a large oscillatory current as this, but if a frame aerial of four turns is employed, the required current is only 84 amperes which is still large but not impracticable. In discussing the application of a loop transmitter to direction finding, we may first consider the vertically polarized field to be the only one radiated. The horizontally polarized field must be discussed later when the possible errors are considered.

The rotating beacon

88. A rotating beacon consists of a loop transmitter, its exact location being notified to those concerned. The loop is usually rotated at a speed of one revolution per minute, this rate being maintained with a high degree of accuracy by means of a special timing mechanism. The figure-of-eight pattern of vertically polarized radiation therefore rotates in space at the same rate.

CHAPTER XVI.—PARAS. 89-91

A special signal, called the north indicating signal, is emitted when the minimum of the diagram is exactly upon the meridian through the transmitter. The method of obtaining a bearing is as follows. The only apparatus required is an ordinary, (i.e. non-directional) W/T receiver, and a stop watch. An operator, listening to the signal, starts the stop watch exactly on receipt of the north indicating signal, and stops it when the figure-of-eight minimum is observed. The number of seconds indicated by the watch, when multiplied by six, gives the true bearing in degrees from north. To avoid the difficulty of correct observation of the north marking signal, when the observer is due north or south of the beacon, an east marking signal is also transmitted. This may be used in exactly the same manner as the north marking signal, but 90° must be subtracted from the calculated bearing.

89. The advantages of this method of direction-finding are :—

- (i) It can be used by any aircraft fitted with a W/T receiver.
- (ii) Any number of aircraft can obtain bearings simultaneously.

The disadvantages being

- (iii) At least 30 seconds are required for each bearing. In this connection it should be noted that it is generally possible to make two observations during each rotation of the loop, giving the bearing and its reciprocal.
- (iv) Since in practice it is desirable to take the mean of a number of observations, the process is rather slow.
- (v) The loop transmitter is subject to vertical effect, which may cause the two minima to be less than 180° apart.
- (vi) The bearing may be in error owing to the reception of abnormally polarized waves.

90. The general design of a rotating beacon transmitter is shown in fig. 39. The whole of the transmitter is fitted inside the loop itself, in order to avoid the necessity for feeding a large radio-frequency current by means of brushes and slip rings. Instead, only the comparatively small anode and filament currents are so carried. The oscillatory circuit is a parallel-feed Hartley. The components, and particularly the frame tuning condenser, must be carefully positioned within the frame in order to reduce the vertical effect to a minimum. The remainder of the vertical component is attributable to unequal currents in the vertical members of the loop, *i.e.* to an unequal distribution of capacitance. It is reduced to a negligible quantity by careful design, particular care being necessary to ensure exact verticality during rotation, which is more difficult than would appear superficially.

91. Let us now briefly consider the horizontally polarized radiation from the loop, assuming the lower horizontal member to be at ground level and the ground to be a perfect conductor. The polar diagrams are then exactly the same as those of a horizontal hertzian doublet of length d , situated at a height h above the ground. The exact diagrams are easily calculated by methods explained in Chapter XIV, but for the present purpose it is sufficient to say that to all intents and purposes the radiation is directed upwards, approximately in the form of an obtuse, elliptical cone. A receiver situated at ground level is therefore unaffected by the direct radiation from the horizontal members, but will be affected by the wave reflected at the ionosphere if its electric field vector has a component parallel to any portion of the receiving aerial. Since the polarization of the wave may be rotated in passing through the ionosphere it is obvious that interference may occur between the direct vertically polarized wave and the reflected wave, so that false minima (or fading which may be mistaken for minima) are likely to be observed. It can be proved that if the ionosphere has exactly the same effect on radiation from all directions, the signals received on a vertical aerial from a loop transmitter will be of the same strength and subject to the same polarization error as would be observed if the signal were transmitted by the vertical aerial with the same radiated power and frequency, and received on the loop aerial. This proposition may have an important application in the siting of a beacon transmitter, *e.g.* the suitability of a suggested site may be tested by locating a portable D/F receiver on the site and observing the systematic errors in bearing of a number of known transmitters. The practical utility of this expedient is still under investigation.

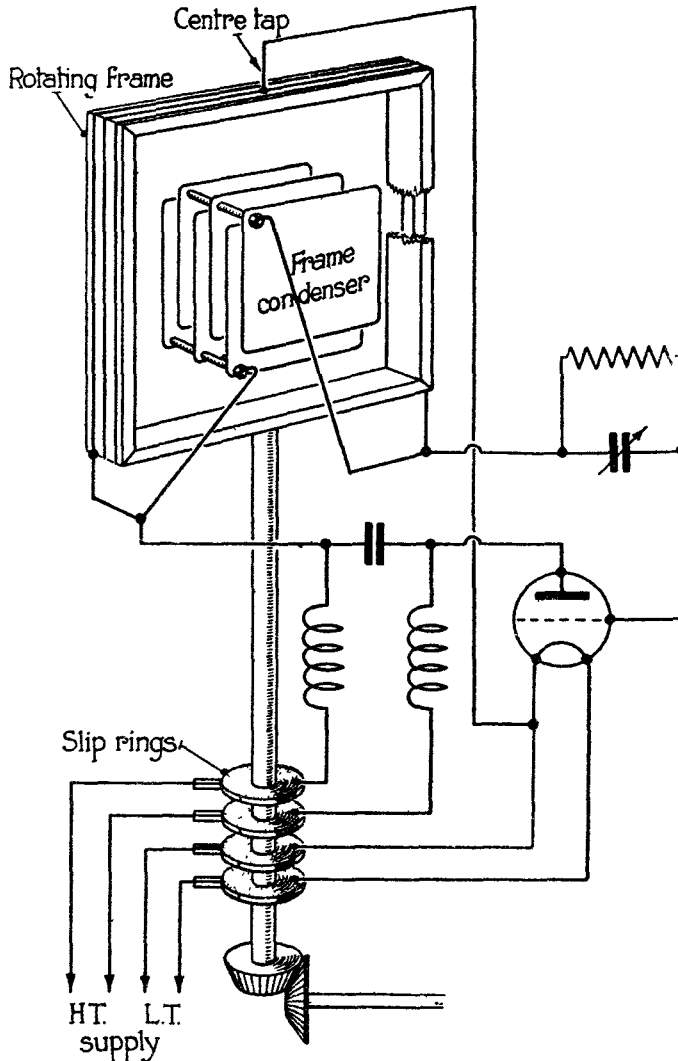


FIG. 39, CHAP. XVI.—Rotating beacon transmitter.

Course or equi-signal beacons

92. This form of beacon employs two directional aerials placed at right angles to each other. The directional aerials may take the form of large loops, say, 100 metres wide and 20 metres high, or, alternatively, an "Adcock" form may be employed. In this case the aerials consist of four vertical aerials supported by towers, 40 metres or so in height, which are placed at the corners of a square of, e.g., 100 metres side and fed from a suitable transmitter by concentric feeder lines. Each pair of diagonally opposite verticals forms one directional aerial system; the two aerials of each pair carry currents in anti-phase so that they function in exactly the same manner as the two sides of a loop aerial, giving a figure-of-eight polar diagram. The "Adcock" arrangement reduces the polarization error to a considerable degree. The principle of operation is shown in fig. 40. The radiation from the two aerials is of equal intensity along

CHAPTER XVI.—PARAS. 93-94

the lines O B, O C, O X, O Y, but elsewhere one of the waves is stronger than the other. Assuming that means are provided by which the two radiations may be distinguished, an aircraft can follow a course along the lines of equal intensity O B, O C, O X, O Y. In the aural-reception equi-signal beacon two opposite letters of the morse code are transmitted by the respective aerials, e.g. "a" by the E W aerial and "n" by the N S aerial. These symbols are interlocked so that the "elements" of one letter fall in the "spaces" of the other. When the aircraft is flying along a line of equal signal intensity, both letters are received and merge into a continuous sound, but if the aircraft is off the desired course to one side "n" is louder than "a", and vice versa.

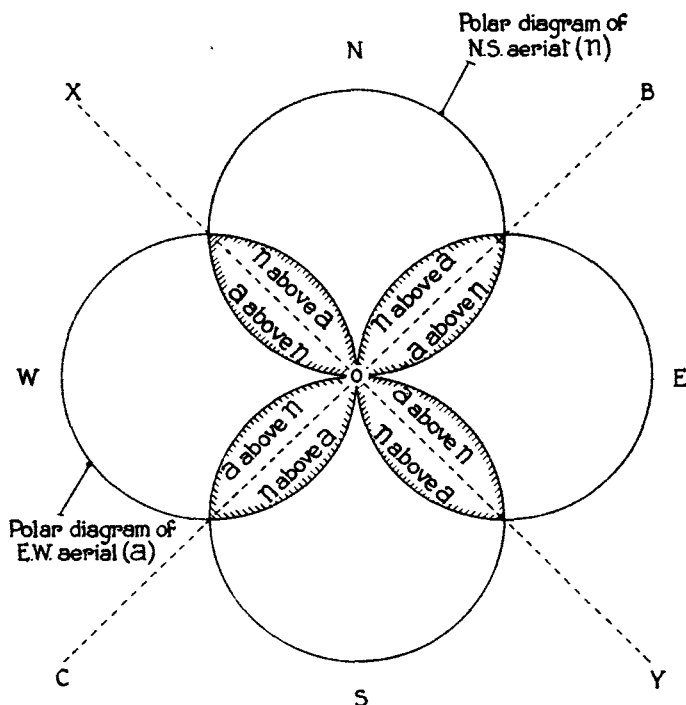


FIG. 40, CHAP. XVI.—Courses laid down by aural-type beacon.

93. Although the aerial systems are fixed in position, the double figure-of-eight radiation pattern can be rotated in azimuth in order that the routes may be marked along any convenient lines of bearing. In the aural-type beacon this is accomplished by coupling the aerial systems to the transmitter by means of a radio-goniometer, consisting of two primary and two secondary windings. The primaries are crossed at right angles, as are also the secondaries, and the angle between the two sets of coils can be varied by relative rotation, so that each primary, in conjunction with its secondary winding and associated aerial system, is in effect equivalent to a loop aerial which is capable of rotation. The coupling between the primary windings and the transmitter proper is by means of a link circuit, the keying being performed by switching from one primary to the other.

94. In the foregoing, it was assumed that the desired routes were perpendicular to each other although this may not be desirable. It is, however, possible to effect a considerable change from the 90° relationship by control of the relative phasing and magnitude of current in the two verticals comprising each aerial of an Adcock-type system, or by the use of additional vertical radiators in the case of the loop-aerial system. Only a few examples need be given to illustrate the possibilities, loop aerials being dealt with as the underlying

theory is simpler than in the case of Adcock aerials. In fig. 41a it is desired to lay down one route in the direction 330° – 150° and a second in the direction 210° – 30° . As these lines are symmetrical about the N S and E W bearings, the loops will have their respective maxima in the latter directions. Let O a b c represent one of the loops of the figure-of-eight pattern produced by the E W loop, the position line 330° – 150° intersecting it at a. Bisect O a. The bisector intersects the line N O at P. With P as centre describe a circle of radius O P, this circle will represent one-half of the figure-of-eight pattern which must be produced by the N S aerial to mark the desired route. This pattern can be produced by reducing the loop current in the N S aerial as compared with that in the E W aerial. In fig. 41b the desired routes are not symmetrically placed with reference to the N S, E W axes. The radiation patterns are produced by the introduction of "in-phase vertical" in the E W aerial, and swung into the required directions by means of the coupling radiogoniometer.

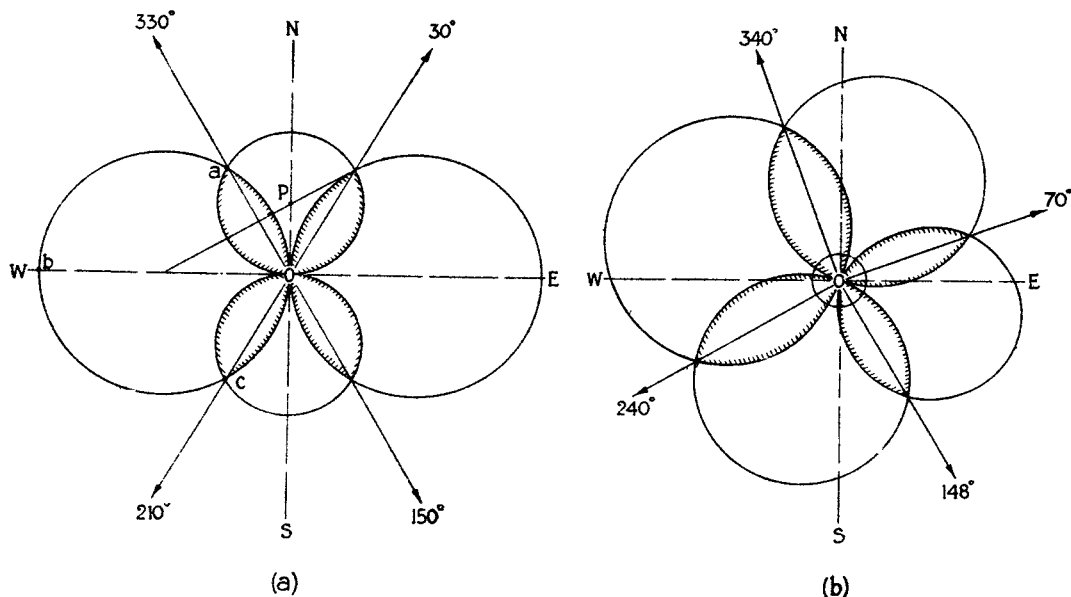


FIG. 41, CHAP. XVI.—Control of courses laid down by equi-signal beacon.

95. In certain instances, the aural-type beacon transmits interlocked "dots" to port and "dashes" to starboard of the marked route, instead of the letters "a" and "n," the dashes being about five times the duration of the dots. It is then possible to design a receiver which is capable of giving simultaneous aural and visual indications. The visual indicator is fed from the secondary of an iron-core transformer, the primary being connected in parallel with the telephones. The secondary winding supplies a voltage to the grid and filament of a triode rectifier. In the anode circuit the rectified current flows through the moving coil of a special form of centre zero instrument. The peculiar feature of this instrument is that it is most sensitive in the region of zero deflection: it is very heavily damped and therefore sluggish in response to a variation of current. The pointer is deflected to the left during the reception of dots and to the right during the reception of dashes, this effect being achieved solely by virtue of the unequal sensitivity and variation of damping at various points on the scale.

96. In the visual type beacon the two radiations are continuously modulated at different audio-frequency rates, e.g. 65 and 86.7 cycles per second. The receiver passes its rectified output to the coil of a small frequency meter of the tuned reed type, carrying only two reeds. When

CHAPTER XVI.—PARAS. 97-99

flying along the equi-signal path both reeds vibrate with equal amplitude. Alternatively, the movement of each reed is caused to induce a small alternating E.M.F. in a suitably placed coil, the resulting current being passed through a metal rectifier and supplied to one winding of a differentially wound, centre-zero moving-coil microammeter. When the amplitude of vibration of both reeds is the same, i.e. when the aircraft is on the desired course, both windings carry equal current and the pointer is not deflected. Off course, the pointer is deflected to port or starboard according to which reed has the greater amplitude, the dial being so marked as to indicate the necessary alteration in course.

97. The routes laid down by the visual type beacon depend upon the depth of modulation and upon the relative phasing of the carrier currents in the two directional aerials. In the aural type these are never simultaneously energized and the question of relative phase does not arise. Suppose the carrier currents in a visual type beacon to be exactly in phase; the sideband energy radiated by each aerial possesses the usual figure-of-eight characteristic, but the carrier energies combine to give a large figure-of-eight rotated through 45° , and as a result only two routes are marked instead of four. If however the carrier currents are 90° out of phase, and the depth of modulation is the same on each carrier, the resulting routes are four in number and are symmetrically disposed. An increase in the depth of modulation of one radiation with respect to the other has the same effect as an increase of aerial current in the aural type beacon. By variation of the relative phasing and depth of modulation, the four routes can be laid down in any direction likely to be required.

Blind approach systems

98. The operation of effecting a landing on an aerodrome under conditions of low or zero visibility may be conveniently divided into three stages :—

- (i) The aircraft must be brought to a point within a few miles of the landing ground.
- (ii) It must then be guided towards the latter on a safe approach track, and finally across the aerodrome itself in a direction suitable for landing, with due regard to the wind.
- (iii) Suitable indications must be given to the aeroplane regarding its distance from the aerodrome during approach, and its height above ground.

Under the first heading, ordinary navigational methods, assisted when necessary by directional wireless, will meet the case. The operations (ii) and (iii) are accomplished by different methods in the several systems of blind landing aids which have been proposed. In the development of any such system, it is essential to take into consideration the available frequency allocations; in Europe, no medium-frequency allocation is possible owing to the demands of the broadcasting services, and the 30-35 Mc/s band is the only one available. The blind approach system developed by the C. Lorenz Company in Germany, which operates entirely on frequencies in this region, has therefore been extensively adopted in Europe for commercial purposes and may be taken as a typical example.

99. The principle of the method is as follows. Referring to fig. 42 an approach path is laid down by an equi-signal beacon which indicates both aurally and visually if the aeroplane is to port or to starboard of the correct line of approach. The equi-signal zone subtends an angle of about 4° , and gives an aural range of about 20 miles, at an altitude of 1,500 feet. Situated upon the line of approach near the actual landing ground are two marker beacons, the inner beacon being about 300 yards outside the aerodrome and the outer beacon some 3,000 yards. The aeroplane is fitted with suitable apparatus for the reception of signals from both main and marker beacons. The bearing upon which the approach track is laid must be known to the pilot; when flying in the equi-signal zone of the main beacon, the pilot receives a steady signal in the telephone receivers, and the pointer of the visual indicator remains steady at zero. When the aeroplane is to the left of the approach path, a succession of dots is received, while if to the right the received signal consists of a series of dashes. Corresponding indications are also given on the visual indicator.

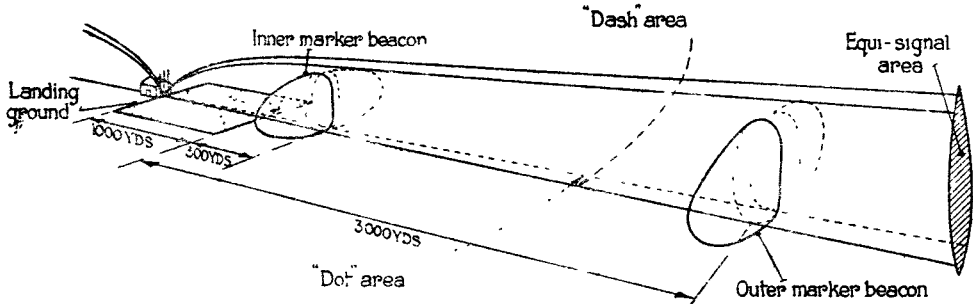


FIG. 42, CHAP. XVI.—Lorenz blind approach system.

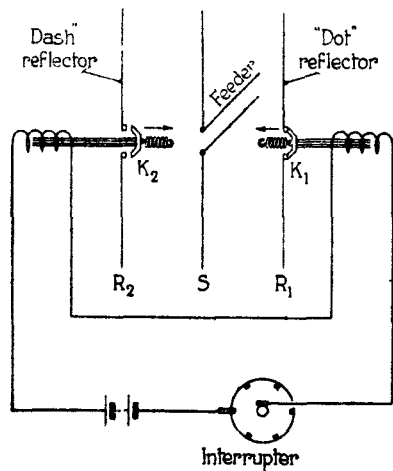
100. The main beacon signals are of the M.C.W. type, the standard modulation frequency being 1,150 cycles per second. On passing over the outer marker beacon, the pilot will hear, superimposed upon the steady note of the main beacon, a deep note of 700 cycles per second, while a very sharp indication of the exact position of the outer beacon is given by the flashing of a neon lamp. Similarly, when the inner marker beacon is approached, a high-pitched note (1,700 cycles per second) is heard and its exact location indicated by another neon lamp. The pilot is now aware that there are no obstacles to his flight in the final section of the landing path. Although not immediately apparent from the foregoing it must be emphasized that the master instrument in the whole manoeuvre is the directional gyro. The actual procedure is to pick up the main beacon at a range of about 20 miles and an altitude of say 2,000 feet. When the equi-signal zone is located, the aeroplane is turned quickly to the known bearing of the approach track, and held steady on this course by the directional gyro. The centre of the approach track may now be maintained with considerable accuracy by observation or either the visual or aural equi-signal indications—a combination of both methods being most effective. Height may be lowered to about 1,000 feet after crossing the outer marker and to below 200 feet after crossing the inner marker, the final stage being of course to touch down on the aerodrome.

101. No mention has been made of the method of navigation in the vertical plane. It has been proposed to employ a locus of constant field strength (i.e. the lower edge of a vertical polar diagram in the equi-signal zone) as a glide path, but although this idea is practicable under certain conditions it is not so generally. Two difficulties arise, namely (i) the shape of the vertical polar diagram does not correspond exactly with the natural glide path of any existing aeroplane. This might be overcome by automatic correction of the intensity of the received signal, (ii) even if the shape of the vertical polar diagram could be correctly adjusted for a given set of conditions, day to day or even hour to hour adjustment might be necessary, because the polar diagram depends to a great extent upon the state of the ground. The present practice therefore is to employ a sensitive Kollsman altimeter for vertical navigation, the barometric pressure at the aerodrome level being given to the pilot by R/T immediately before the landing.

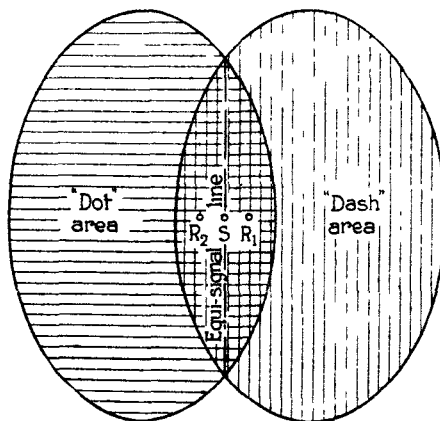
Lorenz system—equipment

102. The manner in which the required equi-signal zone is achieved is illustrated in fig. 43. The aerial system of the main beacon consists of a vertical dipole, which is energized by the main transmitter, and if used alone, the horizontal polar diagram would approximate to a circle. Two reflector dipoles are placed one on each side of the main energized dipole, and are alternately keyed in the manner shown in fig. 43a. When the switch K_1 is closed, the reflector R_1 is resonant with the energized dipole S , but the switch K_2 is broken and the reflector R_2 is completely out of resonance. Under these conditions the polar diagram is that shown by the horizontally shaded area, in fig. 43b. When the switch K_1 is open, K_2 is closed and only the reflector R_2 is operative, the polar diagram being then as shown by the vertically shaded area. The equi-signal zone of the alternately energized aerial-reflector system will therefore be somewhat

CHAPTER XVI.—PARA. 102



(a)



(b)

FIG. 43, CHAP. XVI.—Method of keying main beacon aerial and resultant horizontal polar diagram.

as shown by the overlapping areas. The energized dipole is fed from a crystal-controlled M.C.W. transmitter having a power input of 500 watts, the frequency being in the vicinity of 33 Mc/s, while the marker beacons are supplied by 5-watt transmitters and operate on a higher frequency, e.g. 38 Mc/s. As already stated the radiation from the marker beacons is modulated at 700 cycles per second (outer) and 1,700 cycles per second (inner). The radiating systems of

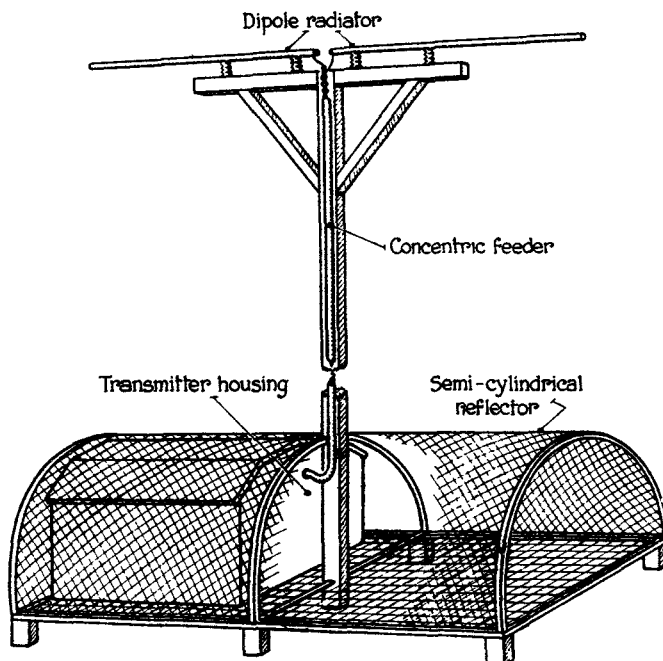


FIG. 44, CHAP. XVI.—Marker beacon.

both marker beacons are of similar design, and the main features are shown in fig. 44. The aerial itself is a horizontal dipole ; the transmitter is situated immediately underneath it, and is enclosed in a screen of expanded metal which is semi-cylindrical in shape. The highest point in the screen is approximately one half-wavelength below the dipole, and the screen, operating as a reflector, gives the peculiarly shaped polar diagrams shown in fig. 45, which are desirable for the function of the marker beacons.

103. The receiving apparatus for the main beacon consists of a simple amplifier having one R.F. stage, with automatic gain control, a detector and one A.F. stage, while a tuned circuit

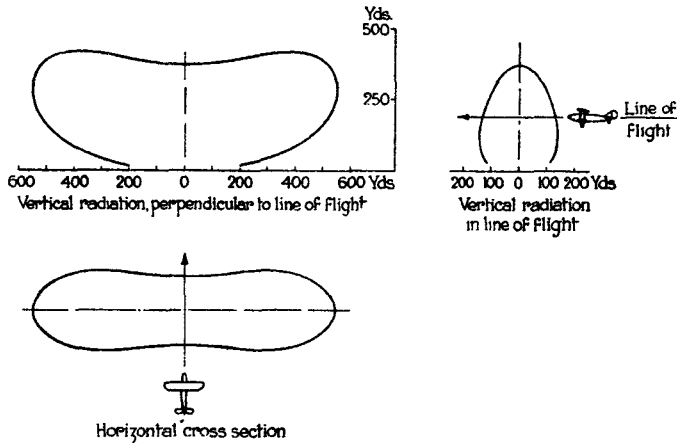


FIG. 45, CHAP XVI.—Polar diagrams of marker beacon.

with a simple detector serves as a receiver for the marker beacon signals, the rectified current from these being fed into the A.F. stage of the main receiver. The output of the latter, during passage through the field of a marker beacon, contains both main and marker modulation frequencies. The output is connected through to suitable filter circuits which supply the appropriate frequencies to the various indicating instruments. The receiving aerial equipment consists of a vertical dipole about $\cdot 9$ metres in height for the reception of the equi-signal beacon, and a horizontal dipole about 1 metre in length for the marker beacons, the latter aerial being generally mounted beneath the fuselage.

APPENDIX A

TABLES

- I. Specific resistance and temperature coefficient.
- II. Properties of insulating materials.
- III. Properties of metals.
- IV. Annealed copper wire, data.
- V. Conversion of British and Metric units.
- VI. Practical and absolute units.
- VII. Mechanical and electrical equivalents.
- VIII. Skin effect factor (F) and proximity factor (G).
- IX. Space factor (J).
- X. Parallel wire factor (w).
- XI. Turns per square inch of winding space.
- XII.
- XIII.
- XIV.
- XV.
- XVI.

APPENDIX A TABLES

TABLE I.—Specific resistance and temperature co-efficient

Material.	Specific resistance at 59° F. = 15° C.			Temperature coefficient.		Resistance compared with copper (approx.)
	Microhms per centimetre cube.	Microhms per inch cube.	Ohms per mil-foot.	Per degree Cent.	Per degree Fahr.	
Metals :—						
Silver, annealed	1·557	0·613	9·36	·00377	0·210	0·92
Copper, pure	1·66	0·654	10·0	·0041	0·23	0·98
Copper, annealed	1·692	0·666	10·18	·00401	0·223	1·0
Copper, hard drawn	1·724	0·679	10·37	·00393	0·218	1·02
Aluminium	2·78	1·09	16·70	·0040	0·22	1·65
Tungsten	4·8	1·9	29·0	·0047	0·26	2·84
Zinc	6·0	2·4	36·0	·0038	0·21	3·55
Tin	11·0	4·3	66·0	·0040	0·22	6·5
Platinum	11·54	4·54	69·4	·0036	0·20	6·82
Nickel (commercial)	11·8	4·6	71·0	·0027	0·15	7·0
Iron telegraph wire.. ..	12·2	4·8	73·4	·0057	0·32	7·2
Lead	21·7	8·55	130·0	·0039	0·22	12·8
Mercury	95·4	37·56	573·0	·0009	0·05	56·0
Alloys :—						
Platinum silver (a)	24·2	9·53	145	·0003	0·015	14·3
German silver (b)	27	10·6	160	·0004	0·02	16·0
German silver (c)	30	11·8	180	·0004	0·02	17·7
Platinoid (d)	42	16·5	250	·0002	0·01	24·8
Manganin (e)	44	17·3	260	Negligible		26·0
Eureka (f)	49	19·3	295	Negligible		28·0
Carbon :—						
Graphite	3,000	1,200	18,000	—0·0005	—0·03	—
Gas retort and lamp filament	4,000	1,600	24,000	—0·0005	—0·03	—
	7,000	2,800	42,000			

(a) 67 per cent. platinum, 33 per cent. silver.

(b) 62 per cent. copper, 15 per cent. nickel, 22 per cent. zinc.

(c) 50 per cent. copper, 30 per cent. nickel, 20 per cent. zinc.

(d) German silver (b) + 1 per cent. tungsten.

(e) 84 per cent. copper, 12 per cent. manganin, 4 per cent. nickel.

(f) 60 per cent. copper, 40 per cent. nickel.

APPENDIX A TABLES

TABLE II.—Properties of insulating materials

Material.	Specific resistance ohms per centimetre cube.	Dielectric strength volts per mil.	Permittivity.
Asbestos	16×10^4	100 to 150	—
Bakelite	10^5 to 10^7	200 to 800	4 to 8
Celluloid	7×10^4	300 to 500	1.2 to 2.6
Cotton, varnished	4×10^8	270 to 350	3
Ebonite	10^9 to 10^{12}	500 to 1,000	2.5 to 3.5
Empire Cloth	4×10^8	500 to 1,500	4 to 6
Fibre, red	100 to 5,000	150 to 300	—
Glass	10^7 — 10^9	200 to 250	5 to 9
Guttapercha	4×10^8	200 to 500	3 to 4.5
Marble	10^8 to 10^5	160	8.3
Mica	10^{10} to 10^{11}	500 to 1,500	5 to 8
Micanite	5×10^8	400 to 1,000	—
Paper, dry	10^5	150 to 300	1.5 to 2.5
Paper, impregnated	10^8	500 to 1,200	2.5 to 4.0
Paraffin wax	4×10^{12}	300	2.0 to 2.5
Porcelain	10^6 to 10^9	150 to 300	4 to 7
Press palm	10^4	200 to 250	—
Quartz	2×10^8	—	2.5 to 4.5
Resin	5×10^{10}	280	2.6
Rubber, pure	Up to 10^{14}	500 to 700	2.2 to 2.4
Rubber, vulcanised	5×10^9	400 to 550	3.0 to 3.5
Shellac	5×10^9	40	3.0 to 3.5
Slate	7.5×10^8	30	6 to 7
Sulphur	10^9 to 10^{11}	—	2.5 to 3.5
Wood	10^5	10 to 20	2.5 to 5.0
Glycerine	—	—	39
Paraffin oil	10^8	220	4.7

Note.—The properties of insulating materials vary so much with temperature, moisture content, and method of manufacture that the above values can only be regarded as approximate.

TABLE III.—Properties of metals

Material.	Specific heat.	Melting point degrees Cent.	Specific gravity.	Weight lb. per cubic in.
Aluminium	.2122	600	2.6	.093
Copper	.0933	1,054	8.8	.32
Iron	.1124	1,600	7.8	.27
Lead	.0315	326	11.3	.41
Mercury	.0385	—40	13.6	.49
Platinum	.0323	1,775	21.5	.78
Silver	.0570	954	10.5	.38
Tin	.0559	230	7.3	.26
Zinc	.0935	412	7.2	.26

TABLE IV.—Annealed copper wire

s.w.g.	Diameter.		Sectional area.		Lb. per 1,000 yards.	Ohms per 1,000 yards (15° C.)
	inch.	m.m.	Sq. inch.	Sq. m.m.		
0	.324	8.23	.08245	53.19	953.4	.2909
1	.300	7.62	.07069	45.60	817.6	.3393
2	.276	7.01	.05983	38.60	692.0	.4009
3	.252	6.401	.04988	32.18	576.7	.4809
4	.232	5.893	.04227	27.27	488.8	.567
5	.212	5.385	.03530	22.77	408.2	.679
6	.192	4.877	.02895	18.68	334.7	.828
7	.176	4.470	.02433	15.70	281.3	.985
8	.160	4.064	.02011	12.97	232.5	1.192
9	.144	3.658	.01629	10.51	188.4	1.471
10	.128	3.251	.01287	8.303	148.8	1.862
11	.116	2.946	.01057	6.819	122.2	2.268
12	.104	2.642	.008495	5.480	98.24	2.821
13	.092	2.337	.006648	4.289	76.88	3.605
14	.080	2.032	.005027	3.243	58.13	4.768
15	.072	1.829	.004072	2.627	47.09	5.89
16	.064	1.626	.003217	2.075	37.20	7.45
17	.056	1.422	.002463	1.589	28.48	9.73
18	.048	1.219	.001810	1.168	20.93	13.24
19	.040	1.016	.001257	.8109	14.53	19.07
20	.036	.9144	.001018	.6567	11.77	23.54
21	.032	.8128	.000804	.5188	9.301	29.80
22	.028	.7112	.000616	.3973	7.120	38.92
23	.024	.6096	.000452	.2919	5.233	53.0
24	.022	.5588	.000380	.2453	4.392	63.0
25	.020	.5080	.000314	.2027	3.633	76.3
26	.018	.4572	.000254	.1642	2.942	94.2
27	.0164	.4166	.000211	.1363	2.44	113.4
28	.0148	.3759	.000172	.1110	1.989	139.3
29	.0136	.3454	.000145	.09372	1.676	165.0
30	.0124	.3149	.000121	.07791	1.399	198.4
31	.0116	.2946	.000106	.06818	1.222	226.8
32	.0108	.2743	.0000916	.05910	1.059	261.6
33	.0100	.2540	.0000785	.05067	.9083	305.1
34	.0092	.2337	.0000665	.04289	.7688	360.5
35	.0084	.2134	.0000554	.03575	.6406	432.4
36	.0076	.1930	.0000454	.02927	.5249	528
37	.0068	.1727	.0000363	.02343	.4197	660
38	.0060	.1524	.0000283	.01824	.3272	848
39	.0052	.1320	.0000213	.01370	.2454	1130
40	.0048	.1220	.0000181	.01170	.2098	1320
41	.0044	.1118	.0000152	.00985	.1760	1580
42	.004	.1030	.0000126	.00833	.1450	1910
43	.0036	.0910	.0000102	.00650	.1180	2350
44	.0032	.0082	.00000805	.00529	.0940	2970
45	.0028	.0711	.00000615	.00399	.0710	3890
46	.0024	.0609	.00000451	.00259	.0540	5300
47	.002	.0507	.00000314	.00203	.0360	7630
48	.0016	.0406	.00000201	.00130	.0230	11900
49	.0012	.0304	.00000113	.00073	.0013	21100
50	.001	.0254	.00000078	.00051	.0091	30510

APPENDIX A TABLES

TABLE V.—Conversion of British and Metric units

Units.	C.G.S. to British.	British to C.G.S.
Length {	1 cm. = 0·3937 in. 1 metre = 3·281 ft. 1 kilometre = 0·6214 mile.	1 in. = 2·54 cms. 1 ft. = 30·48 cms. 1 mile = 1609 metres.
Area {	1 sq. cm. = 0·155 sq. in. 1 sq. metre = 10·76 sq. ft.	1 sq. in. = 6·452 sq. cms. 1 sq. ft. = 0·0929 sq. metres.
Volume {	1 cu. cm. = 0·061 cu. in. 1 cu. metre = 35·32 cu. ft.	1 cu. in. = 16·39 cu. cms. 1 cu. ft. = 0·0283 cu. metres.
Mass {	1 gm. = 0·0353 oz. 1 kilogram. = 2·205 lb.	1 oz. = 28·35 gms. 1 lb. = 0·4536 kilogram.

TABLE VI.—Values in absolute units

Electrical quantity.	Practical Unit.	Value in absolute E.M. Units.
Current	Ampere	10 ⁻¹
Electromotive force, P.D.	Volt	10 ⁸
Resistance	Ohm	10 ⁹
Quantity	Coulomb	10 ⁻¹
Energy	Joule	10 ⁷
Power	Watt	10 ⁷
Power	Kilowatt	10 ¹⁰
Capacitance	Farad	10 ⁻⁹
Inductance	Henry	10 ⁹

TABLE VII.—Mechanical and Electrical work and power equivalents

N.B.—1 B.Th.U. = amount of heat required to raise 1 lb. of water through 1° F.

Units.	Equivalent values in other units.
1 B.Th.U.	1,055 watt seconds. 778 ft.-lb. 0·252 calorie (kg.d). 108 kilogram-metres. 0·000291 kilowatt hour. 0·000388 H.P. hour. 0·0000667 lb. coal (oxidised). 0·00087 lb. water evaporated at 212° F.
1 B.Th.U. per square foot per min. ..	0·121 watts per square inch. 0·0174 kilowatt. 0·0232 H.P.
1 ft.-lb.	1·36 joule. 0·1383 kilogram-metre. 0·000000377 kilowatt hour. 0·00128 B.Th.U. 0·0000005 H P. hour.

APPENDIX A TABLES

TABLE VII.—Mechanical and Electrical work and power equivalents—*contd.*

Units.	Equivalent values in other units.
1 lb. water evaporated at 212° F. ..	0.33 kilowatt hour. 0.44 H.P. hour. 1.140 B.Th.U. 124,200 kilogram-metres. 1,219,000 joules. 887,800 ft.-lb.
1 H.P.	746 watts. 0.746 kilowatts. 33,000 ft.-lb. per minute. 550 ft.-lb. per second. 2,548 B.Th.U. per hour. 42.5 B.Th.U. per minute. 0.71 B.Th.U. per second. 2.25 lb. water evaporated per hour at 212° F.
1 H.P. hour	0.746 kilowatt hour. 1,980,000 ft.-lb. 2,548 B.Th.U. 273,740 kilogram-metres. 2.25 lb. water evaporated at 212° F. 17.2 lb. water raised from 62° to 212° F.
1 Watt	1 joule per second. 0.00134 H.P. 0.001 kilowatt. 3.415 B.Th.U. per hour. 44.25 B.Th.U. per minute. 0.73 ft.-lb. per second. 0.003 lb. of water evaporated per hour.
1 Kilowatt	1,000 watts. 1.34 H.P. 2,655,200 ft.-lb. per hour. 44,253 ft.-lb. per minute. 737.5 ft. lb. per second. 3,415 B.Th.U. per hour. 57.0 B.Th.U. per minute. 0.95 B.Th.U. per second. 3 lb. water evaporated per hour at 212° F.
1 Kilowatt hour	1,000 watt hours. 1.34 H.P. hours. 2,655,200 ft.-lb. 3,600,000 joules. 3,415 B.Th.U. 366,848 kilogram-metres. 3 lb. water evaporated at 212° F. 22.8 lb. water raised from 62° to 212° F.
1 Kilogram-metre	7.23 ft.-lb. 0.0000366 H.P. hour. 0.0000272 kilowatt hour. 0.0092 B.Th.U.
1 Joule	1 watt second. 0.0000278 kilowatt hour. 0.102 kilogram-metre. 0.000947 B.Th.U. 0.73 ft.-lb.

APPENDIX A TABLES

TABLE VIII.—Skin effect and proximity factors

F = Skin effect factor. G = proximity factor. d = diameter of wire in cm. ρ = specific resistance in E.M. Units. For copper, $\rho = 1700$ approx., and $z = .1078 d\sqrt{f}$.

z	F	G	z	F	G	z	F	G	z	F	G
0.0	negligible	$\left. \begin{array}{l} z^4 \\ 64 \end{array} \right\}$	2.5	0.175	0.2949	5.0	1.043	0.755	10.0	2.799	1.641
0.1			2.6	0.201	0.3184	5.2	1.114	0.790	11.0	3.151	1.818
0.2			2.7	0.228	0.3412	5.4	1.184	0.826	12.0	3.504	1.995
0.3			2.8	0.256	0.3632	5.6	1.254	0.861	13.0	3.856	2.171
0.4			2.9	0.286	0.3844	5.8	1.324	0.896	14.0	4.209	2.348
0.5			3.0	0.318	0.4049	6.0	1.394	0.932	15.0	4.562	2.525
		0.00097									
0.6	0.001	0.00202	3.1	0.351	0.4247	6.2	1.463	0.967	16.0	4.915	2.702
0.7	0.001	0.00373	3.2	0.385	0.4439	6.4	1.533	1.003	17.0	5.268	2.879
0.8	0.002	0.00632	3.3	0.420	0.4626	6.6	1.603	1.038	18.0	5.621	3.056
0.9	0.003	0.01006	3.4	0.456	0.4807	6.8	1.673	1.073	19.0	5.974	3.233
1.0	0.005	0.01519	3.5	0.492	0.4987	7.0	1.743	1.109	20.0	6.328	3.409
1.1	0.008	0.02196	3.6	0.529	0.5160	7.2	1.813	1.144	21.0	6.681	3.586
1.2	0.011	0.03059	3.7	0.566	0.5333	7.4	1.884	1.180	22.0	7.034	3.763
1.3	0.015	0.04127	3.8	0.603	0.5503	7.6	1.954	1.216	23.0	7.388	3.940
1.4	0.020	0.0541	3.9	0.640	0.5673	7.8	2.024	1.251	24.0	7.741	4.117
1.5	0.026	0.0691	4.0	0.678	0.5842	8.0	2.094	1.287	25.0	8.094	4.294
1.6	0.033	0.0863	4.1	0.715	0.601	8.2	2.165	1.322	30.0	9.86	5.177
1.7	0.042	0.1055	4.2	0.752	0.618	8.4	2.235	1.357	40.0	13.40	6.946
1.8	0.052	0.1265	4.3	0.789	0.635	8.6	2.306	1.393	50.0	16.93	8.713
1.9	0.064	0.1489	4.4	0.826	0.652	8.8	2.376	1.428	60.0	20.46	10.48
2.0	0.078	0.1724	4.5	0.863	0.669	9.0	2.446	1.464	70.0	24.00	12.25
2.1	0.094	0.1967	4.6	0.899	0.686	9.2	2.517	1.499	80.0	27.54	14.02
2.2	0.111	0.2214	4.7	0.935	0.703	9.4	2.587	1.534	90.0	31.07	15.78
2.3	0.131	0.2462	4.8	0.971	0.720	9.6	2.658	1.570	100.0	34.61	17.55
2.4	0.152	0.2708	4.9	1.007	0.738	9.8	2.728	1.605	>100	$\frac{\sqrt{2z}-3}{4}$	$\frac{\sqrt{2z}-1}{8}$
2.5	0.175	0.2949	5.0	1.043	0.755	10.0	2.799	1.641			

TABLE IX.—Space factor, J

For use in the formula $R_{ac} = R_{dc} \left(1 + F + \frac{u G c^2}{1 - J d^2} \right)$ (Chapter V).

(i) Currents in adjacent wires in the same direction

z	0	1	2	3	4	5	>>5
J	0.042	0.033	-0.056	-0.152	-0.176	-0.185	-0.250

(ii) Currents in adjacent wires in opposite directions.

z	0	1	2	3	4	5	>>5
J	0.042	0.053	0.169	0.348	0.466	0.53	0.750

APPENDIX A TABLES

TABLE X.—Factor μ . (Chapter V)

μ depends upon the number of wires running parallel, as under.

No. of wires	2	4	6	8	10
μ	1.0	1.8	2.16	2.37	2.51
No. of wires	12	16	24	32	∞
μ	2.61	2.74	2.91	3.0	$\frac{\pi^2}{3}$

The above values are applicable to single-layer solenoids or pancake coils in which the winding space is small compared with the radius of the coil.

For single-layer solenoids of many turns, well spaced, the following Table may be used.

D = diameter of coil.			l = length of coil.			
$\frac{l}{D}$	0	0.2	0.4	0.6	0.8	1.0
μ	3.29	3.63	4.06	4.50	4.93	5.29
$\frac{l}{D}$	2	4	6	8	10	∞
μ	6.58	7.74	8.38	8.73	8.9	9.87

APPENDIX A TABLES

TABLE XI.—*Number of turns per square inch of winding space*

s.w.g.	Double cotton covered.	Single cotton covered.	Double silk covered.	Single silk covered.	Enamel. Covered.
10	49	54	57	58	59
11	59	65	69	71	72
12	72	80	85	87	89
13	89	100	107	111	114
14	113	129	140	145	150
15	135	156	171	178	185
16	173	198	213	223	234
17	216	252	274	287	304
18	297	343	377	400	403
19	400	472	528	567	587
20	472	567	641	692	720
21	567	692	793	865	910
22	692	865	1,110	1,110	1,173
23	865	1,110	1,320	1,480	1,575
24	977	1,280	1,600	1,770	1,880
25	1,110	1,480	1,890	2,110	2,250
26	1,280	1,740	2,270	2,560	2,770
27	1,450	1,980	2,670	3,010	3,300
28	1,630	2,310	3,160	3,650	4,100
29	1,810	2,600	3,600	4,580	4,770
30	1,990	2,950	4,500	5,180	5,660

APPENDIX B
MATHEMATICAL NOTES

- I. Trigonometrical formulae.
- II. Hyperbolic functions.
- III. Harmonic analysis.
- IV.
- V.
- VI.
- VII.
- VIII.

APPENDIX B

MATHEMATICAL NOTES

I. TRIGONOMETRICAL FORMULAE

1. The complement of an angle is the angle which must be added to it to make a right angle. Thus the complement of 60° is 30° , and of $\frac{\pi}{6}$ radians is $\frac{\pi}{3}$ radians. The trigonometrical ratios of an angle and its complement are related as follows.

$$\begin{array}{ll} \sin (90 - A) = \cos A & \cos (90 - A) = \sin A \\ \tan (90 - A) = \cot A & \cot (90 - A) = \tan A \\ \operatorname{cosec} (90 - A) = \sec A & \sec (90 - A) = \operatorname{cosec} A. \end{array}$$

These formulae are valid for all values of A , whether greater or less than 90° .

2. The supplement of an angle is the angle which must be added to it to make two right angles. Thus the supplement of 60° is 120° , and of $\frac{\pi}{4}$ radians is $\frac{3}{4}\pi$ radians. The trigonometrical ratios of an angle and its supplement are related as follows.

$$\begin{array}{ll} \sin (180 - A) = \sin A & \cos (180 - A) = -\cos A \\ \tan (180 - A) = -\tan A & \cot (180 - A) = -\cot A \\ \operatorname{cosec} (180 - A) = \operatorname{cosec} A & \sec (180 - A) = -\sec A. \end{array}$$

3. The following are known as the Addition Formulae.

$$\begin{array}{l} \sin (A + B) = \sin A \cos B + \sin B \cos A \\ \sin (A - B) = \sin A \cos B - \sin B \cos A \\ \cos (A + B) = \cos A \cos B - \sin A \sin B \\ \cos (A - B) = \cos A \cos B + \sin A \sin B. \end{array}$$

4. By the addition and subtraction of various pairs of the above the following relations are obtained

$$\begin{array}{l} \sin (A + B) + \sin (A - B) = 2 \sin A \cos B \\ \sin (A + B) - \sin (A - B) = 2 \cos A \sin B \\ \cos (A - B) + \cos (A + B) = 2 \cos A \cos B \\ \cos (A - B) - \cos (A + B) = 2 \sin A \sin B. \end{array}$$

5. By putting $(A + B) = P$, $(A - B) = Q$,

$$\begin{array}{l} \sin P + \sin Q = 2 \sin \frac{P+Q}{2} \cos \frac{P-Q}{2} \\ \sin P - \sin Q = 2 \cos \frac{P+Q}{2} \sin \frac{P-Q}{2} \\ \cos Q + \cos P = 2 \cos \frac{P+Q}{2} \cos \frac{P-Q}{2} \\ \cos Q - \cos P = 2 \sin \frac{P+Q}{2} \sin \frac{P-Q}{2}. \end{array}$$

6. The following properties of the tangent are also important.

$$\begin{array}{l} \tan (A + B) = \frac{\sin (A + B)}{\cos (A + B)} = \frac{\tan A + \tan B}{1 - \tan A \tan B} \\ \tan (A - B) = \frac{\tan A - \tan B}{1 + \tan A \tan B}. \end{array}$$

APPENDIX B

MATHEMATICAL NOTES

7. Many other useful relations may be deduced from the Addition Formulae; an frequently used are :—

$$\sin 2A = 2 \sin A \cos A$$

$$\cos 2A = \cos^2 A - \sin^2 A$$

$$= 1 - 2 \sin^2 A$$

$$= 2 \cos^2 A - 1$$

$$\tan 2A = \frac{2 \tan A}{1 - \tan^2 A}$$

$$\tan A = \frac{2 \tan \frac{A}{2}}{1 - \tan^2 \frac{A}{2}}$$

$$\frac{1 - \cos A}{1 + \cos A} = \tan^2 A$$

$$\sin 3A = 3 \sin A - 4 \sin^3 A$$

$$\cos 3A = 4 \cos^3 A - 3 \cos A$$

$$\tan 3A = \frac{3 \tan A - \tan^3 A}{1 - 3 \tan^2 A}$$

APPENDIX B

MATHEMATICAL NOTES

II.—HYPERBOLIC FUNCTIONS

1. In dealing with transmission lines, both for radio and audio frequencies, considerable use is made of hyperbolic functions. The direct hyperbolic functions of a quantity x are its hyperbolic sine, hyperbolic cosine, hyperbolic tangent, etc., and these have corresponding inverse functions. The hyperbolic functions are so called because they are connected with the geometry of the rectangular hyperbola in a manner resembling the relation between the circular functions, sine, cosine, etc., and the angle subtended by a circular arc, but this geometrical relation is of no interest from the engineering point of view. They are most easily understood if treated merely as combinations of exponential functions.

2. The direct hyperbolic functions are written $\sinh x$, $\cosh x$, etc., and are defined as follows:—

$$\sinh x = \frac{e^x - e^{-x}}{2}$$

$$\cosh x = \frac{e^x + e^{-x}}{2}$$

$$\tanh x = \frac{e^x - e^{-x}}{e^x + e^{-x}} = \frac{\sinh x}{\cosh x}$$

$$\coth x = \frac{\cosh x}{\sinh x} = \frac{1}{\tanh x}$$

$$\operatorname{cosech} x = \frac{1}{\sinh x}$$

$$\operatorname{sech} x = \frac{1}{\cosh x}$$

("Sinh" is pronounced as "shin," "tanh" as "than," and "sech" as "shec").

3. The circular functions $\sin \varphi$, $\cos \varphi$, may be expressed in the forms of infinite series, thus

$$\sin \varphi = \varphi - \frac{\varphi^3}{3!} + \frac{\varphi^5}{5!} - \dots$$

$$\cos \varphi = 1 - \frac{\varphi^2}{2!} + \frac{\varphi^4}{4!} - \dots$$

Similarly, the hyperbolic functions may be represented by infinite series, for example,

$$\sinh x = x + \frac{x^3}{3!} + \frac{x^5}{5!} + \dots$$

$$\cosh x = 1 + \frac{x^2}{2!} + \frac{x^4}{4!} + \dots$$

4. The values of $\cosh x$, $\sinh x$, and $\tanh x$, from $x = 0$ to $x = 2.5$, are shown graphically in fig. 1. It is obvious from the formulae already given that $\sinh 0 = 0$, and $\cosh 0 = 1$. When x exceeds about 3, e^{-x} becomes small compared with e^x . Both $\sinh x$ and $\cosh x$ then approach more closely to the value $\frac{e^x}{2}$ as x increases. The functions e^x , e^{-x} , are given in many books of Tables, and may also be evaluated with sufficient accuracy for most purposes by means of a Log-log slide rule. The following relations between circular and hyperbolic functions are of primary importance.

5. In Chapter V it is shown that

$$e^{j\varphi} = \cos \varphi + j \sin \varphi$$

$$e^{-j\varphi} = \cos \varphi - j \sin \varphi$$

Hence

$$\cos \varphi = \frac{e^{j\varphi} + e^{-j\varphi}}{2} = \cosh j\varphi$$

$$\sin \varphi = \frac{e^{j\varphi} - e^{-j\varphi}}{2j} = \frac{\sinh j\varphi}{j}$$

APPENDIX B

MATHEMATICAL NOTES

6. If instead of φ we write $j\varphi$,

$$\cos j\varphi = \frac{e^{j(j\varphi)} + e^{-j(j\varphi)}}{2} = \frac{e^{-\varphi} + e^{+\varphi}}{2},$$

i.e.

$$\cos j\varphi = \cosh \varphi$$

Also

$$\sin j\varphi = \frac{e^{j(j\varphi)} - e^{-j(j\varphi)}}{2j} = \frac{e^{-\varphi} - e^{+\varphi}}{2j}$$

But

$$\frac{e^{-\varphi} - e^{+\varphi}}{2j} = j \frac{e^{\varphi} - e^{-\varphi}}{2}$$

Hence

$$\sin j\varphi = j \sinh \varphi$$

Again,

$$\tan j\varphi = \frac{\sin j\varphi}{\cos j\varphi} = j \frac{\sinh \varphi}{\cosh \varphi} = j \tanh \varphi$$

$$\cot j\varphi = \frac{\cos j\varphi}{\sin j\varphi} = \frac{\cosh \varphi}{j \sinh \varphi} = -j \coth \varphi.$$

7. Collecting and extending the above results,

$\sinh j\varphi = j \sinh \varphi$	$\sin j\varphi = j \sinh \varphi$
$\cosh j\varphi = \cosh \varphi$	$\cos j\varphi = \cosh \varphi$
$\tanh j\varphi = j \tanh \varphi$	$\tan j\varphi = j \tanh \varphi$
$\coth j\varphi = -j \coth \varphi$	$\cot j\varphi = -j \coth \varphi$
$\operatorname{sech} j\varphi = \sec \varphi$	$\sec j\varphi = \operatorname{sech} \varphi$
$\operatorname{cosech} j\varphi = -j \operatorname{cosec} \varphi$	$\operatorname{cosec} j\varphi = -j \operatorname{cosech} \varphi.$

8. The Addition Formulae for hyperbolic functions are as below :—

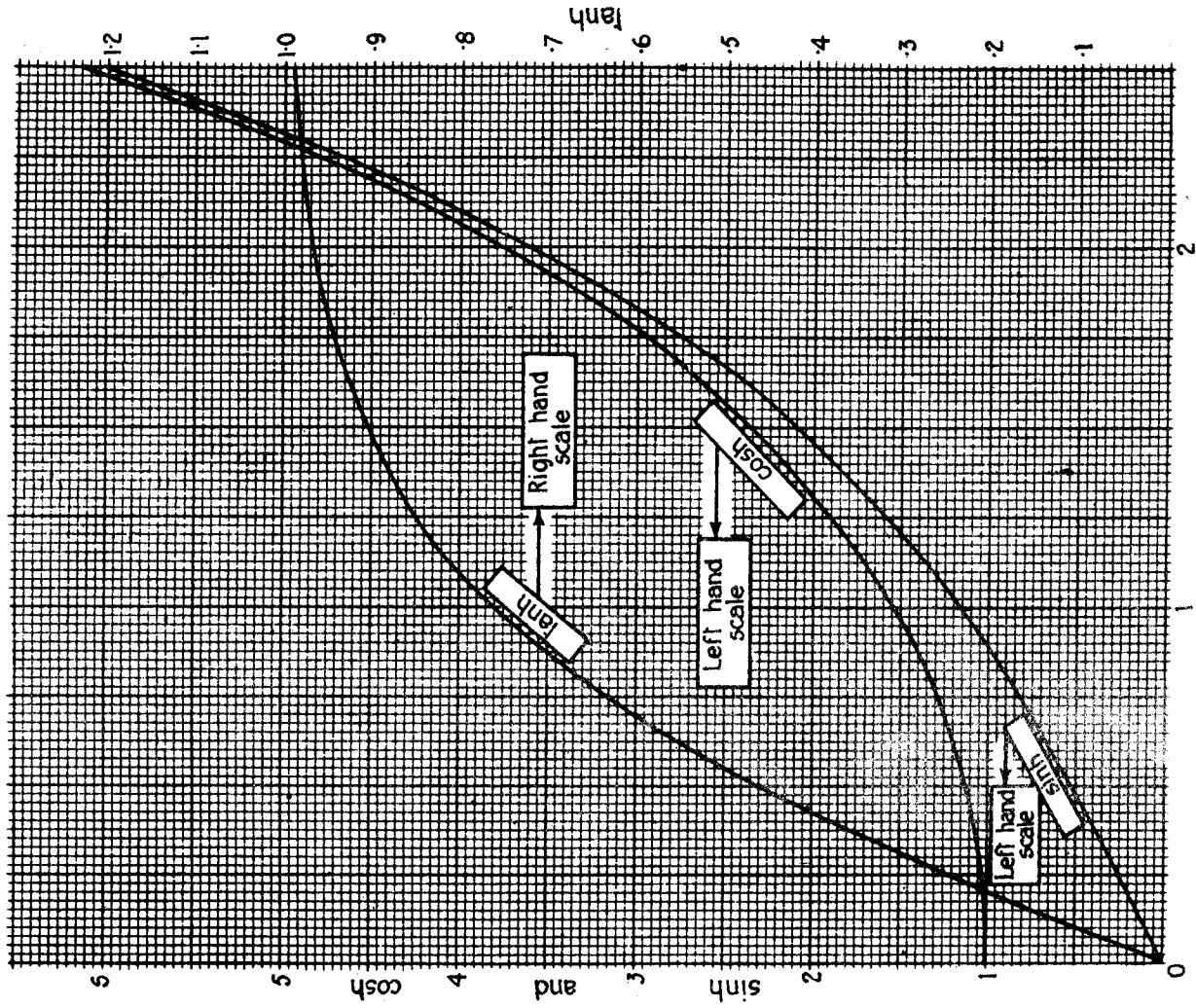
$$\begin{aligned} \sinh (A + B) &= \sinh A \cosh B + \cosh A \sinh B \\ \sinh (A - B) &= \sinh A \cosh B - \cosh A \sinh B \\ \cosh (A + B) &= \cosh A \cosh B + \sinh A \sinh B \\ \cosh (A - B) &= \cosh A \cosh B - \sinh A \sinh B. \end{aligned}$$

9. By addition and subtraction of various pairs of the above, the following relations are obtained

$$\begin{aligned} \sinh (A + B) + \sinh (A - B) &= 2 \sinh A \cosh B \\ \sinh (A + B) - \sinh (A - B) &= 2 \cosh A \sinh B \\ \cosh (A + B) + \cosh (A - B) &= 2 \cosh A \cosh B \\ \cosh (A + B) - \cosh (A - B) &= 2 \sinh A \sinh B \end{aligned}$$

10. By putting $(A + B) = P$, $(A - B) = Q$,

$$\begin{aligned} \cosh P + \cosh Q &= 2 \cosh \frac{P+Q}{2} \cosh \frac{P-Q}{2} \\ \cosh P - \cosh Q &= 2 \sinh \frac{P+Q}{2} \sinh \frac{P-Q}{2} \\ \sinh P + \sinh Q &= 2 \sinh \frac{P+Q}{2} \cosh \frac{P-Q}{2} \\ \sinh P - \sinh Q &= 2 \cosh \frac{P+Q}{2} \sinh \frac{P-Q}{2}. \end{aligned}$$



HYPERBOLIC FUNCTIONS

FIG.1
APPENDIX B

APPENDIX B

MATHEMATICAL NOTES

11. Other useful relations are :—

$$\begin{aligned}\tanh (A+B) &= \frac{\sinh A \cosh B + \cosh A \sinh B}{\cosh A \cosh B + \sinh A \sinh B} \\ &= \frac{\tanh A + \tanh B}{1 + \tanh A \tanh B} \\ \tanh (A-B) &= \frac{\tanh A - \tanh B}{1 - \tanh A \tanh B} \\ \cosh^2 A - \sinh^2 A &= 1 \\ \sinh 2A &= 2 \sinh A \cosh A \\ \cosh 2A &= \cosh^2 A + \sinh^2 A \\ &= 1 + 2 \sinh^2 A \\ &= 2 \cosh^2 A - 1 \\ \tanh 2A &= \frac{2 \tanh A}{1 + \tanh^2 A} = \frac{2 \coth A}{1 + \coth^2 A} = \frac{2}{\coth A + \tanh A}.\end{aligned}$$

12. Expressions such as $\sinh (A + jB)$, which are met with in the theory of telephonic transmission, are dealt with in the following manner :—

$$\begin{aligned}\sinh (A + jB) &= \sinh A \cosh jB + \cosh A \sinh jB \\ \cosh jB &= \cos B \\ \sinh jB &= j \sin B\end{aligned}$$

therefore

$$\sinh (A + jB) = \sinh A \cos B + j \cosh A \sin B$$

Similarly

$$\begin{aligned}\cosh (A + jB) &= \cosh A \cos B + j \sinh A \sin B \\ \sinh (A - jB) &= \sinh A \cos B - j \cosh A \sin B \\ \cosh (A - jB) &= \cosh A \cos B - j \sinh A \sin B\end{aligned}$$

and

$$\begin{aligned}\tanh (A + jB) &= \frac{\sinh (A + jB)}{\cosh (A + jB)} \\ \coth (A + jB) &= \frac{\cosh (A + jB)}{\sinh (A + jB)}\end{aligned}$$

Thus, the hyperbolic functions of complex numbers are also complex.

Example.—Evaluate $\sinh (A + jB)$ when $A = 0.7$ and $B = 1$ radian = 57 degrees approx.
From tables of circular and hyperbolic functions

$$\begin{aligned}\sin B &= 0.8387 & \cos B &= 0.5446 \\ \sinh A &= 0.7586 & \cosh A &= 0.6044 \\ \sinh (A + jB) &= 0.7586 \times 0.5446 + j 0.6044 \times 0.8387 \\ &= 0.4131 + j 0.5069.\end{aligned}$$

APPENDIX B

MATHEMATICAL NOTES

III.—HARMONIC ANALYSIS

1. When the graphical representation of one or more cycles of a complex periodic wave is given, the analysis of the waveform into its constituent harmonics, up to and including the sixth, is easily performed by a procedure known as the twelve-ordinate method. The mathematical theory will not be given; it is sufficient to say that the process has been reduced to simple arithmetic by means of the schedules given below. As a concrete example, the waveform of fig. 2 will be analysed.

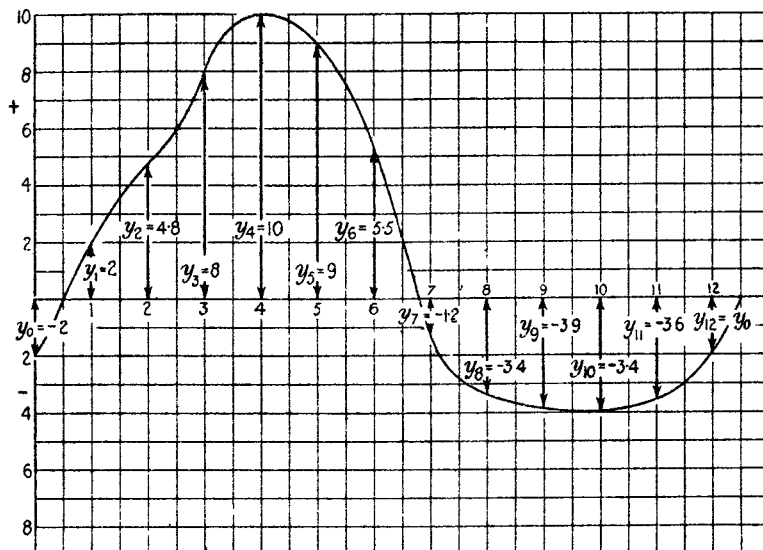


FIG. 2, APPENDIX B.—Analysis of complex wave-form.

2. In the diagram, one complete cycle of 360° has been divided into 12 equal parts, and the corresponding ordinates erected. Their values are then tabulated as below:—

y_0	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}	y_{11}
-2.0	2.0	4.8	8.0	10.0	9.0	5.5	-1.2	-3.4	-3.9	-4.0	-3.6

The computation is performed as follows. First, enter the values of the ordinates upon the following schedule (a), finding the sums (Σ) and differences (Δ) as indicated; $s_1 = y_1 + y_{11}$, $d_1 = y_1 - y_{11}$ and so on.

Schedule (a)

	y_0	y_1	y_2	y_3	y_4	y_5	y_6
		y_{11}	y_{10}	y_9	y_8	y_7	
(Σ)	..	..	s_0	s_1	s_2	s_3	s_4
(Δ)	..	..		d_1	d_2	d_3	d_4

3. Substituting actual values of $y_0, y_1, \dots, y_6$ from the table, Schedule (a) becomes

-2	2.0	4.8	8.0	10.0	9.0	5.5
-3.6	-4.0	-3.9	-3.4	-1.2		

(Σ)	..	..	-2	-1.6	0.8	4.1	6.6	7.8	5.5
(Δ)	..	..		5.6	8.8	11.9	13.4	10.2	

i.e.

$$s_0 = -2.0, s_1 = -1.6, s_2 = 0.8, s_3 = 4.1, s_4 = 6.6, s_5 = 7.8, s_6 = 5.5.$$

$$d_1 = 5.6, d_2 = 8.8, d_3 = 11.9, d_4 = 13.4, d_5 = 10.2.$$

APPENDIX B

MATHEMATICAL NOTES

4. Next, find the sums and differences in the schedules (b) and (c) .—

Schedule (b)

			s_0	s_1	s_2	s_3
			s_4	s_5	s_6	
(Σ)	..	..	S_0	S_1	S_2	S_3
(Δ)	..	..	D_0	D_1	D_2	

Schedule (c)

			d_1	d_2	d_3
			d_4	d_5	
(Σ)	..	..	S_4	S_5	S_6
(Δ)	..	..	D_3	D_4	

5. Inserting numerical values from schedule (a), schedule (b) becomes

			-2	-1.6	0.8	4.1
			5.5	7.8	6.6	
(Σ)	..	.	3.5	6.2	7.4	4.1
(Δ)	..	..	-7.5	-9.4	-5.8	

and schedule (c),

			5.6	8.8	11.9
			10.2	13.4	
(Σ)	..	..	15.8	22.2	11.9
(Δ)	..	..	-4.6	-4.6	

i.e. $S_0 = 3.5$, $S_1 = 6.2$, $S_2 = 7.4$, $S_3 = 4.1$, $S_4 = 15.8$

$S_5 = 22.2$, $S_6 = 11.9$

$D_0 = -7.5$, $D_1 = -9.4$, $D_2 = -5.8$, $D_3 = -4.6$, $D_4 = -4.6$

6. Finally complete the schedules (d) and (e) below

Schedule (d)

		S_0	S_1
		S_2	S_3
(Σ)	..	S_7	S_8

Schedule (e)

		S_4	D_0
		S_5	D_1
(Δ)	..	D_5	D_6

7. Inserting numerical values from previous results

Schedule (d) becomes

		3.5	6.2
		7.4	4.1
(Σ)	..	10.9	10.3

Schedule (e) becomes

		15.8	-7.5
		11.9	-5.8
(Δ)	..	3.9	-1.7

i.e. $S_7 = 10.9$, $S_8 = 10.3$, $D_5 = 3.9$, $D_6 = -1.7$.

The coefficients $S_0, \dots, S_8, D_0, \dots, D_6$ are the quantities actually required for the analysis.

APPENDIX B

MATHEMATICAL NOTES

8. The equation of the wave-form is

$$y = A_0 + A_1 \cos \omega t + A_2 \cos 2 \omega t + A_3 \cos 3 \omega t + A_4 \cos 4 \omega t + A_5 \cos 5 \omega t + A_6 \cos 6 \omega t + B_1 \sin \omega t + B_2 \sin 2 \omega t + B_3 \sin 3 \omega t + B_4 \sin 4 \omega t + B_5 \sin 5 \omega t$$

where

$$\begin{aligned} A_0 &= \frac{S_7 + S_8}{12} ; \\ A_1 &= \frac{D_0 + \frac{\sqrt{3}}{2}D_1 + \frac{1}{2}D_2}{6} ; \quad B_1 = \frac{\frac{1}{2}S_4 + \frac{\sqrt{3}}{2}S_5 + S_6}{6} \\ A_2 &= \frac{S_0 + \frac{1}{2}(S_1 - S_2) - S_3}{6} ; \quad B_2 = \frac{\frac{\sqrt{3}}{2}(D_3 + D_4)}{6} \\ A_3 &= \frac{D_3}{6} ; \quad B_3 = \frac{D_6}{6} \\ A_4 &= \frac{S_0 - \frac{1}{2}(S_1 + S_2) + S_3}{6} ; \quad B_4 = \frac{\frac{\sqrt{3}}{2}(D_3 - D_4)}{6} \\ A_5 &= \frac{D_3 - \frac{\sqrt{3}}{2}D_1 + \frac{1}{2}D_2}{6} ; \quad B_5 = \frac{\frac{1}{2}S_4 - \frac{\sqrt{3}}{2}S_5 + S_6}{6} \\ A_6 &= \frac{S_7 - S_8}{12} \end{aligned}$$

9. The term A_0 is a constant displacement, e.g. the mean anode current in the case of a valve circuit. The amplitude H_1 of the fundamental is $\sqrt{A_1^2 + B_1^2}$ and of the n^{th} harmonic ($n \geq 6$) is $\sqrt{A_n^2 + B_n^2}$. The instantaneous value of the n^{th} harmonic is

$$h_n = \sqrt{A_n^2 + B_n^2} \sin(n \omega t + \varphi_n)$$

where

$$\varphi_n = \tan^{-1} \frac{A_n}{B_n}$$

10. Returning to the analysis of fig. 2,

$$\begin{aligned} A_0 &= \frac{10.9 + 10.3}{12} = 1.77 \\ A_1 &= \frac{-7.5 + (0.866 \times -9.4) - 2.9}{6} = \frac{-18.5}{6} = -3.1. \\ B_1 &= \frac{(\frac{1}{2} \times 15.8) + (0.866 \times 22.2) + 11.9}{6} = 6.5. \\ h_1 &= 7.2 \sin \left(\omega t - \tan^{-1} \frac{3.1}{6.5} \right) \\ A_2 &= \frac{3.5 + \frac{1}{2}(6.2 - 7.4) - 4.1}{6} = \frac{-1.2}{6} = -0.2 \\ B_2 &= \frac{.866(-4.6 - 4.6)}{6} = -1.33 \\ A_3 &= \frac{-1.7}{6} = -0.28 \\ B_3 &= \frac{3.9}{6} = 0.65 \\ A_4 &= \frac{3.5 - \frac{1}{2}(6.2 + 7.4) + 4.1}{6} = \frac{0.8}{6} = 0.133 \\ B_4 &= \frac{.866(-4.6 + 4.6)}{6} = 0. \\ A_5 &= \frac{-7.5 + (0.866 \times 0.94) - 2.9}{6} = \frac{-2.26}{6} = -0.38 \\ B_5 &= \frac{7.9 - (0.866 \times 22.2) + 11.9}{6} = \frac{0.4}{6} = 0.067 \\ A_6 &= \frac{10.9 - 10.3}{12} = 0.05 \end{aligned}$$

APPENDIX B

MATHEMATICAL NOTES

11. The following checks should be applied to ensure arithmetical accuracy. For the A terms:—

$$A_0 + A_1 + A_2 + A_3 + A_4 + A_5 + A_6 = y_0.$$

i.e.

$$1.77 - 3.1 - 0.2 - 0.28 + 0.133 - 0.38 + 0.05 = y_0 \\ \text{or } y_0 = 2.007.$$

This is well within the accuracy with which the value of y_0 can be determined from the graph.

The accuracy of the B terms is checked by

$$(B_1 + B_2) + \sqrt{3} (B_3 + B_4) + 2 B_5 = y_1 - y_{11}$$

i.e.

$$(6.5 + 0.067) + \sqrt{3} (-1.33) + (2 \times 0.65) = y_1 - y_{11} \\ \text{or } y_1 - y_{11} = 2 - (-3.567) \\ = 5.567$$

which is again satisfactory. From the table, $y_1 - y_{11} = 5.6$.

12. The amplitudes of the various harmonics are

$$H_1 = \sqrt{A_1^2 + B_1^2} = 7.2$$

$$H_2 = \sqrt{A_2^2 + B_2^2} = 1.35$$

$$H_3 = \sqrt{A_3^2 + B_3^2} = 0.707$$

$$H_4 = \sqrt{A_4^2 + B_4^2} = 0.133$$

$$H_5 = \sqrt{A_5^2 + B_5^2} = 0.38$$

$$H_6 = \sqrt{A_6^2 + B_6^2} = 0.05$$

The percentage of second harmonic is $\frac{H_2}{H_1} \times 100 = 18.7$,

of third harmonic, $\frac{H_3}{H_1} \times 100 = 9.8$,

and so on.

INDEX

A		Ch.	Para.	C		Ch.	Para.
A.C. Circuits		5	12	Calibration D/F		16	33
Acceptor Circuits		5	38	Capacitance		1	53
A.C. Values Instant Peak RMS. ..		5	1	AC		5	22
Activation		8	11	Bridge		13	41
Aerials. Adcock System		16	66	Inter-Electrode		8	31
Arrays		15	135	Measurement		13	28
Arrays Receiving		15	149	Stray		11	10
Counterpoise		7	45	C.G.S. System		11	8
Directive		15	15	Characteristic-Curves		8	19
Doublet and Dipole	7 52	15	10	Chirp or Ripple		9	41
Frame		16	7	Circuit Calculations		1	31
High Frequency		7	56	Coastal Refraction	16 23	16	88
Inverted		7	64	Coercive Force		2	23
Loop	16 3	16	30	Commutation		4	16
Practical		7	40	Compounds		1	5
Receiving		10	1	Constants Valve		8	18
Reflector		15	33	Condenser		1	54
Theory		15	1	Charge and Discharge		1	59
Alternators		4	3	Energy Stored		1	57
Amplification				Condensers		1	73
Audio Frequency		11	16	Conductors		1	42
Factor		8	21	Conductance. Conductivity	1 26	14	29
HF. and VHF.		11	77	Coulomb		1	21
Modulated Waves		11	116	Coupling. Coupled Circuits		6	23
Power		11	33	Choke Capacity		11	21
Principles		11	1	Resistance Capacity		11	17
Radio Frequency		11	45	Resonant Circuits		6	1
Regenerative		10	28	Transformer		11	26
Amplifiers. A.F. Power		12	67	Parallel Feed		11	32
Classification		11	2	Tuned Anode Capacity		11	46
Class A.	11 4	12	73	Tuned Transformer		11	48
Class B.		11	106	Counterpoise		7	45
Class C.		11	119	Conversion Angle		16	42
Multi Stage		11	72	Crystal Control		9	58
Paraphase		12	76	Currents		1	12
Receiving		9	13	Direction of Flow		1	19
RF Power	11 103	11	139	Feed. Aerial		7	58
Amplitude		4	2	Cycle		4	2
Ampere		1	21	D			
Swimming Rule		2	6	Damping		7	13
Anode. Dissipation		8	8	Anode Load		10	27
Load		11	59	Grid Current		10	25
Tap		9	27	Dead Space		10	17
Appleton Layer		14	52	Dead Beat		13	2
Armature Reaction		4	28	Decibel		10	57
Windings		4	20	Demoivres Theorem		5	60
Atoms		1	4	Detection AM		10	8
Attenuation		14	44	FM		12	88
Atmosphere		14	49	Dia-Magnetic		2	23
Atmospherics		10	59	Diffraction		14	8
Autodyne		10	32	Dielectric Constant		1	51
A.V.C.		12	66	Strength		1	61
B				Direction Finding		16	1
Bandwidth Fractional		11	3	Aircraft		16	30
Beacons Radio		16	84	Ground Station		16	52
Beat Effect		10	13	Directional Radiation		15	15
Bearings		16	9	Dissipation Line		8	30
Bellini Tosi		16	53	Displacement Effect		16	21
Blind Approach		16	98	Dispersion		14	22
Booster		4	55	Distortion. Amplitude		12	49
B.O.T.U.		1	39	Amplifiers		11	4
Brewster Angle		14	22	Frequency		12	39

	Ch.	Para.
Double Super Heterodyne ..	11	102
Dry Cells	1	25
Dyne	viii	18
Dynamometer	5	6
Dynamic Characteristics ..	8	25

E

Eddy Currents	4	12
Electrical Units	1	21
Electrification	1	11
Electrons	1	6
Electronic Emission	8	1
Electro Dynamic Laws	7	23
Electro-magnetism	2	26
Electro-magnetic Wave	7	26
Electrostatics	1	47
Elements	1	5
EMF.	1	17
Chemical Action	1	23
Alternating	4	1
Induced	2	24
Magnitude	4	25
E.M.U.'s	2	10
E.S.U.'s	1	49
Energy Density	2	36
Kinetic, Potential	1	52
Erg.	viii	18
Errors D/F .. 16 10 16 23 16 72	16	83
Ether	7	20
Evacuation Process	8	11
Exponential Form	5	63

F

Fading	14	59
Faraday's Laws	2	26
Feeders H.F.	7	57
R.F.	15	83
Non Resonant	7	61
Twin, and Concentric	15	101
Ferro-Magnetic	2	13
Field Windings	4	33
Strength	1	50
Figure of Merit Valve	11	42
Flemings L.H. Rule	2	17
R.H. Rule	2	25
Flux, and Density	1	48
Fresnels Equations	14	18
Frequency AC.	4	9
Changers	11	85
Control, Variation	9	47
Modulation	12	79
Multipliers	11	140
R.F.	7	8
R F. Oscillator	11	98
Range E M. Waves	14	1
F.P.S. System	vi	8

G

Gain and Stability	11	70
Gain Control	12	58
Galvanometer	2	9
Gauss	1	11
Generators D.C.	4	15
Goniometer Radio	16	53
Grating Factor	15	16

	Ch.	Para.
Grid Bias	9	24
Automatic	11	134
Current Effect	11	56
Leak and Condenser	10 26	12 55
Current	8	23
Gettering	8	11
Gravity	vi	11

H

Harmonics	5 2 9 23	12 5
Harmonic Distortion	11 38	12 49
Heaviside Layer	14	48
Heterodyne	10	13
Separate	10	16
High-Mu Triode	11	135
Hyperbolic Functions	15	73
Hysteresis	2	23
Horse Power	ix	20
Hygroscopic Substances	1	67

I

I C.W.	9	11
Impedance	5	20
Mutual	5 19	15 42
Surge	15	92
Impulse Excitation	9	21
Inductance	2	30
A.C.	5	15
Inductances	2	37
Induction	2	24
Mutual	2	28
Input Admittance	11	53
Insulating Material's	1	62
Instruments	3 1 5 5	5 28
Interference	12	90
Signal—Noise Ratio	11	100
Superheterodyne	1	11
Ions	1 11	14 27
Ionisation	14	51
Ionosphere	1	9
Isotopic Materials	1	9

J

Joule	ix 19	1 40
-------------	-------	------

K

Kennelly—Heaviside Layer	12	48
Keying	9	43
Key Clicks	9	12
Kilo-volt Ampere	5	35
Kirchoffs Laws	1	31

L

L.C.R. Parallel Combinations	5	43
Lamination	4 12	6 1
Lenz's Law	2	27
Listening Through. Limiter Valve	10	48
Load Line	8	28
Load Impedance	11 7	12 26
Losses. Aerials	7	54
Copper	6	11
Generators D.C.	4	40
Iron	6	16
Motors D.C.	4	54
Transformers	6	9

M	Ch.	Para.
Magnetism and E.M.	2	1
Magnetic Field	2	2
Magnetic Field, Strength	2	10
Electro	7	23
Effect	14	32
Magnetic Flux	2	11
D.P.	2	14
Magneto Motive Force	2	19
Magnification, Circuit	13	33
Voltage	5	41
Matching Devices	15	107
Matter	1	1
Maxwell's Rules	1 37	2 6
M.C.W.	9	26
Meters	3	1
A.C.	5	5
Amps. and Volts	3 6	5 5
Frequency	5	7
Power	3 16	5 28
Ohms	3	13
Metalising	8	11
Microphones	1 46	12 6
Miller Effect	10	27
M H.O.	1	26
Modulation, Amplitude	12	13
Gross	12	61
Frequency	12	79
Indicator	13	42
Inter	12	41
Methods	12	21
Over	11	18
Phase	12	78
Modulated Amplifiers Anode	11	124
Grid	11	129
High-Mu Triode	11	135
Modulated Aerial Current	12	20
Molecules	1	3
Motors D.C.	4	44
Motor Generator	4	54
Starters	4	51
N		
Natural Frequency	7	8
Neon Tube	13	6
Neutralization	9 51 11	136
Neutron	1	10
Note Selector	11	22
Noise in Amplifiers	11	76
Non-ohmic Conductors	1	27
Networks	1	36
Nucleus	1	8
O		
Octantal Error	16	72
Ohms Law	1	26
Oscillations, Mechanical	7	1
Electrical	7	7
Intermittent	9	24
Modulated	12	14
U.H.F.	9	65
Oscillators, Barkhausen—Kurz	9	65
Biassed to cut-off	9	21
Circuits	9	8
Crystal	9	61
Master	9	49
Power and Efficiency	9	16

Oscillators—	Ch.	Para.
Push Pull	9	33
Radio Frequency	11	98
Series/Shunt Feed	9	9
Output Methods	11	43
Power	11	34
Undistorted	11	39
Ozonosphere	14	50
P		
Para-Magnetic	2	23
Parallel Feed	11	32
Peak Voltmeter	13	23
Permeameter	2	21
Permittivity	1	15
P.D.	1	18
Polarity, Magnetic	2	1
Solenoid	2	8
Reversal	4	39
Polarization		
Primary Cells	1	23
Wave	7 32	14 15
Change	14	57
Error	16	60
Positron	1	10
Power Calculations	1	39
Carrier and Sidebands	12 18	12 24
Output	4 26	11 34
Supplies	9	37
Primary Cells	1	23
Propagation	14	1
Along Ground	14	41
Formulae	14	45
High Frequency	14	67
L.F. and V.L.F.	14	61
Protons	1	7
Pulse Coil	9	30
Push Pull	9 33	12 72
Q		
"Q" Circuit Value	5	42
Quiescent Push Pull	12	74
Quadrantal Correction	16	31
R		
Radiation	7	34
Resistance	7 51	15 8
Radio Compass	16	47
Telephony	12	1
Reception	12	45
Reactance, Inductive	5	17
Capacitive	5	23
Reaction	10	28
Control	10	43
Effects	11	55
Negative	10	45
Receivers, C.W.	10 42	11 81
Reception, R.T.	12	45
Resonance, Series	5	36
Parallel	5	48
Coupled Circuits	6	26
Indicators	13	3
Resonant Frequency	5 36	5 48
Rectification	10	4
Anode	10 35	12 51
Diode	10 5	12 57
Full Wave	9	42
Grid	10 21	12 54

Rectification—	Ch.	Para.
Half Wave	9	39
Triode	10 20 12	50
Rectifiers, Valve and Metal	9	43
Reflection	14	7
Coefficient	14 35 15	59
Refraction	14	9
Coastal	16 23 16	83
Rejector Circuit	5	50
Resistance	1	28
A.C.	5	12
Colour Code	1	44
Radio Frequency	13	34
Resistances, Resistors	1	42
Retentivity	2	23
Ring Mains	1	36
Ripple	9	41
R.M.S.	5	1
Rotation Reversal	4	38

S

Safety Devices	9	43
Sal-Ammoniac	1	24
Self Oscillation	9 50 11 65 11	74
Selectivity	12	46
S/F Amplifier	11	99
Separate Heterodyne	10	16
Send Receive Switch	10	47
Sensefinding	16	24
Screened Loop	16	19
Screened Grid	8	32
Side Bends	12	14
Tone	12	38
Sine Curve	4	5
Site Selection D/F	16	73
Skin Effect	5 12	46
Skip Distance	14	67
Solar Influences	14	56
Sound	12	2
Sparkling, Commutators	4	30
Stability and Gain	11	70
Stage Gain	11	18
Standing Waves	7	41
Sunspot Cycle	14	77
Superheterodyne Receiver	11	81
Suppressed Carrier, and Side Band		
Telephony	12	42
Suppression Unit	15	132
Smoothing	9	41
Stratosphere	14	50

T

Temperature Coefficient	1 30	1 42
Telephone Receiver	2	41
Thermionic Emission	8	2
Thyratron	8	12
Tonic Train	9	26
Toroidal Coil	2	20
Torque	4	27
Transformer	6	1
Auto	6	17
Design	6	20
Efficiency	6	16
Load	6	3
Losses	6	9
Output	11	43
R.F. Shielded	16	20

Transmitter C.W.	Ch.	Para.
H.F.	9	63
M.O.	9	52
Modulation	12	27
Tuning	9	14
Transmission, Lines	7 57	15 66
Non Resonant	7	61
Troposphere Tropopause	14	49
Tuning, Aerial	7 48	10 3
Harmonies	13	14
Methods	11	75
Reflector	15	53
Wavemeter	13	11
Fork Control	9	57
Tuned Anode	11 12	11 66

U

U.H.F. Oscillation	9	65
Amplification	11	77

V

Values, R.M.S., etc.	5	1
Valves, Acorn	11	71
Construction	8	10
Class B	12	75
Diode	8 13	10 48
Electrodes	8	4
Gasfilled, Hard, Soft	8	12
Hexode	11	96
Heptode, Pentagrid	11	92
Magnetron	9	67
Mercury Vapour	9	44
Octode	11	94
Pentode	8 37	12 70
Power	12	68
Tetrode	8 32	11 69
Triode	8	15
Triode-Hexode	11	97
Variable-Mu	12	64
Valves in Parallel	9 32	12 69
Push-Pull	9 33	12 72
Valve Volt Meter	13	20
Vectors A.C.	5	8
Operators A.C.	5	54
Algebra	15	25
Velocity Angular	4	6
E.M. Waves	7	19
Phase and Group	14	13
Vertical Effect	16	17
Volt	1	27
Voltage	1	18
Amplification Factor	11	6
Doubling	9	46
Feed, Aerial	7	60
Magnification	5	41
Transformation	4	53

W

Wavelength	7	18
Waves		
Electromagnetic	7	17
Form	4 2	5 2
Forms, Types	7	39
Standing	7	41
Wavemeters	13	1
Wheatstone Bridge	1	38
Wattmeter	5	29